Inventory and Monitoring Program

National Park Service
U.S. Department of the Interior

DRAFT

San Francisco Bay Area Network
Inventory and Monitoring Program

Data Management Plan

David Press
Lead Data Manager, SFAN

1. Introduction

1.1. Purpose and Scope of the San Francisco Bay Area Network Data Management Plan

Collecting natural resource data is our first step toward understanding the ecosystems within our national parks. These ecosystems are evolving, as is our knowledge of them and how they work. We use these "raw" data to analyze, synthesize, and model aspects of ecosystems. In turn, we use our results and interpretations to make decisions about the Park's vital natural resources. Thus, *data* collected by researchers and maintained by the San Francisco Bay Area Network (SFAN) Data Management Plan will become *information* through analyses, syntheses, and modeling.

Information is the common currency among the many and various activities and people involved in stewardship projects throughout our National Park System. These projects may include park planning, creating inventories, short-term and long-term monitoring, restoration, control of invasive species and other species management, fire management, trail and road maintenance, law enforcement, and the communication of natural resource information to the public.

One of the National Park Service's primary goals is to "improve park management through greater reliance on scientific knowledge." Thus, the SFAN Inventory and Monitoring Program will identify, catalog, organize, structure, archive, and provide high-quality natural resource data to institutions, researchers, managers, and administrators. The SFAN Inventory and Monitoring Program is designed to ensure the integrity of these data for many years to come.

1.1.1. Data and Data Management: An Overview

Any good set of data, whether collected last week or 20 years ago, must tell us enough about itself, so that we can reliably preserve and use it. Anyone using these data will need to know as much as possible about how and why these data were collected. Therefore, data sets must be accompanied by a sufficient amount of documentation describing how and why it was collected.

The SFAN Data Management Program cannot simply attend to the tables, fields, and values that make up a data set. It must also provide a process for developing, preserving, and integrating the context that makes data interpretable and valuable. Although this means more time documenting, it leads us to clearer preservation and presentation of data.

We sometimes use the term "data" in a broader way to encompass other products that are generated alongside primary tabular and spatial data. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records.

To meet Inventory and Monitoring (I&M) Program goals, and to ensure adequate context for primary data products, some level of management is needed to ensure their quality and availability. We intend to integrate the manner in which the Network creates, manages, and provides the products of our research and analysis. All phases of data and information processing are integrated, and information about each phase and its processes must be shared through good documentation.

1.1.2. Data Management Prioritization

There are many potential sources of important data and information about the condition of natural resources in our parks. Because the I&M Program focuses on long-term monitoring and natural resource inventories, our first priority should be toward the data and information that we derive from these primary efforts. However, we can easily apply our same standards, procedures, infrastructure, and attitudes about data management to other natural resource data sources.

The highest data management priority should be placed on producing and archiving high-quality, well-documented data originating from the I&M Program. As time and resources permit, we will work toward raising the level of data management for current projects, legacy data, and data originating outside the I&M Program. In addition, high-profile data sets that provide crucial information to park management should be prioritized for data management regardless of funding source.

1.2. Inventory and Monitoring Program Overview

The I&M Program represents a long-term commitment by the National Park Service to assess and document the status and trends of park ecological resources. In 1998, the National Parks Omnibus Management Act established the framework for the I&M Program to fully integrate natural resource monitoring and other scientific activities into the management processes of the National Park System.

The Act charges the Secretary of the Interior to "continually improve the ability of the National Park Service to provide state-of-the-art management, protection, and interpretation of and research on the resources of the National Park System", and to "… assure the full and proper utilization of the results of scientific studies for park management decisions." Section 5934 of the Act requires the Secretary of the Interior to develop a program of "inventory and monitoring of National Park System resources to establish baseline information and to provide information on the long-term trends in the condition of National Park System resources."

To carry out this mission, the National Park Service initiated a service-wide, natural resource Inventory and Monitoring Program encompassing 270 parks with significant natural resources. Ecologically similar parks among the 270 chosen were grouped into 32 networks. Each I&M Network has been tasked with documenting existing park vertebrates and vascular plants, developing a management based, ecological monitoring program with a written plan and protocols, as well as a Data Management Plan that encompasses all aspects of the Network program.

The SFAN includes eight parks with significant natural resources in the central California region. These parks include Eugene O'Neill National Historic Site (EUON), Fort Point National Historic Site (FOPO), Golden Gate NRA (GOGA), John Muir National Historic Site (JOMU), Muir Woods National Monument (MUWO), Pinnacles National Monument (PINN), the Presidio of San Francisco (PRES) and Point Reyes National Seashore (PORE). FOPO, MUWO, and PRES are within the boundaries of and are administered by GOGA, so are included as part of GOGA

for the purposes of this Data Management Plan. I&M projects are not managed locally at EUON and JOMU, but are instead administered by Network staff based at either PORE or GOGA. I&M data management specifications for EUON and JOMU are thus incorporated into other parts of this Data Management Plan.

The SFAN has included two parks that were not identified in the national list of 270 parks with significant natural resources for which the servicewide program was designed. PRES has several areas of significant natural resources, such as Crissy Field, so it was included in the SFAN. The SFAN also included EUON because it is jointly managed with JOMU and is surrounded on three sides by Las Trampas Regional Wilderness Park. Therefore, wildlife may utilize EUON and significant plant communities occur nearby.

The SFAN personnel are headquartered in Building 1063 at Fort Cronkite in the Marin Headlands region of the Golden Gate National Recreation Area. Network personnel based at this office in FY2005 include the Network Coordinator, Inventory Coordinator, Lead Data Manager, Network Ecologist, and a Network funded Biological Technician (Data Miner). With the five-year inventory efforts ending in FY2005, the Inventory Coordinator and Network Data Miner positions will end in mid-2005. Because the Network parks are located far apart from one another, specific I&M project management and oversight primarily occurs at the park level. Decentralizing the SFAN I&M program maintains the continuity and expertise of inventory and monitoring efforts occurring within the Network parks.

1.2.1. SFAN Inventory and Monitoring Program's Long-Term Goals

The I&M Program's success at identifying, cataloging, organizing, structuring, archiving, and providing relevant natural resource information will largely determine the Program's efficacy and image among critics, peers, and advocates.

The I&M Program's long-term goals are to:

- establish natural resource inventory and monitoring standards throughout the National Park system that transcend traditional program, activity, and funding boundaries
- inventory the natural resources and park ecosystems under National Park Service stewardship
- monitor park ecosystems to provide reference points for comparisons with other, altered environments
- integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making
- share National Park Service accomplishments and information with other natural resource organizations and form partnerships for attaining common goals and objectives

To achieve the last two of these goals, we must develop a modern information management infrastructure (e.g., staffing, hardware, software). This infrastructure will include procedures to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers, and others will be entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available to others for management decision-making, research, and education.

The National Park Service is a highly decentralized agency with complex data requirements. The primary audience for many of the products from the I&M Program is at the park level, to provide park managers with the information they need to make better-informed decisions, and to work more effectively with other agencies and individuals for the benefit of park resources. However, certain data are also needed at the regional or national level for a variety of purposes. The National Park Service Advisory Board (2001) has stated that the findings "must be communicated to the public, for it is the broader public that will decide the fate of these resources."

1.3. Data Management Plan Overview

The SFAN Data Management Plan outlines how we intend to implement and maintain a system that will serve the data and information management needs of our Inventory and Monitoring Program. This plan reflects our commitment to the establishment, maintenance, description, accessibility, and long-term availability of high-quality data and information.

The SFAN Data Management Plan describes how the network will:

- support I&M Program objectives
- acquire and process data
- assure data quality
- document, analyze, summarize and disseminate data and information
- maintain nationally developed data management systems
- maintain, store and archive data

1.3.1. Data Management Goals and Objectives

As the basic and most important products of scientific research, data and information represent a valuable, and often, irreplaceable resource. Because field experiments and associated data collection are often time and budget consuming, management of data and information products plays an important role in any scientific program. For long-term ecological monitoring programs, such as the SFAN Inventory and Monitoring Program, retention and documentation of high quality data are the foundation upon which the success of the overall I&M Program rests.

The SFAN approach to data management is to develop a Data Management Plan that is "user friendly" toward a varied audience consisting of park natural resource managers to data managers. This means providing guidance on data management practices at a number of different levels. To facilitate this guidance, we developed the SFAN plan along three basic principles:

1. Keep the plan simple, flexible, and evolving;
2. make the plan useful to all – to park GIS and data management staff, regional technical staff, resource management staff, and cooperating scientists; and,
3. include the data users in the decision making process whenever possible.

1.3.2. Types of Data Covered by the Data Management Plan

This plan covers four major categories of data that are coordinated or managed by the SFAN. These are:

1. **Data managed in servicewide databases.** The SFAN uses three data systems developed by the I&M WASO office. NatureBib is used as a bibliographic tool for cataloging reports, publications, or other documents that relate to natural resources in park units. Dataset Catalog is used to document primarily non-spatial databases or other data assemblages. NPSpecies is used by the network to develop and maintain species lists in network parks, along with associated supporting evidence.

2. **Data developed or acquired directly by the network as a result of inventory, monitoring, or other projects.** This category includes project-related protocols, reports, spatial data, and associated materials such as field notes and photographs provided to SFAN by contractors or developed by SFAN staff. Projects can be short-term (one to two years duration) or long-term (ongoing monitoring).

3. **Data that, while not developed or maintained by SFAN, are used as data sources or provide context to other data sets.** Examples of this category include: GIS data developed by parks, other agencies or organizations; national or international taxonomic or other classification systems; climate or hydrologic data collected by regional or national entities.

These above categories can contain one or more of the following data formats:

- hard-copy documents (e.g., reports, field notes, survey forms, maps, references, administrative documents)
- objects (e.g., specimens, samples, photographs, slides)
- electronic files (e.g., Word files, email, websites, digital images)
- electronic tabular data (e.g., databases, spreadsheets, tables, delimited files)
- spatial data (e.g., shapefiles, coverages, remote-sensing data)

Each of these data formats has specific requirements for ongoing management and maintenance, which we address in this plan.

1.3.3. Revisions

The SFAN data management plan covers I&M Program needs based on the most current information systems technology relevant through 2004. Revisions to this plan and associated data management documents (guidelines and SOPs) will be completed as needed.

1.3.4. Intended Audience for the SFAN Data Management Plan

SFAN developed this information management plan in modular format with the hope of increasing both its readability and usability among a wider audience. This format allows information plan users to easily locate and access subdocuments pertaining to a particular element of information management guidance or process. Individual procedural documents or chapters from within the plan can be provided as standalone documents to park and regional data management staff, cooperators and project leaders. This format also allows for a more complete, technical review and promotes quality through consistent implementation of a process or procedure within the organization, even if there are temporary or permanent personnel changes. Technical experts can be provided with applicable portions of the plan based on their expertise (GIS, Database, Metadata, GPS, etc.).

1.3.5. SFAN Data Management Guidance Documents Associated with this Plan

Forthcoming documents related to the Data Management Plan will be made available on the SFAN website (http://science.nature.nps.gov/im/units/SFAN/) once completed. These documents will be intended as instructions or Standard Operating Procedures for specific data management tasks. SOPs will expand on chapters presented in this Data Management Plan and will address a broad audience. These SOPs will include, but will not be limited to:

- SFAN I&M Program Data Storage and Archiving Procedures
- SFAN I&M Program Guidelines for developing FGDC compliant metadata for both spatial and biological data
- SFAN I&M Program Standards for File Revision Control
- SFAN I&M Program Database Template Specifications and Users Manual
- SFAN I&M Program NPSpecies Data Management Plan
- SFAN I&M Program NatureBib Data Management Plan
- SFAN I&M Program QA/QC procedures for spatial and biological datasets.
- SFAN I&M Program GIS Development and Management Guidelines
- SFAN I&M Program Legacy Database Conversion Procedures

2. Data Management Roles and Responsibilities

Data management is about people and organizations as much as it is about information technology and database theory and application. Nearly every person in an organization manages data and information at some level. Understanding how and why the National Park Service manages its data is a central duty each person has in order to help the organization deliver products and services.

This chapter discusses the Network's programmatic data management roles and responsibilities. Additional details based on this framework are found in Vital Sign monitoring plans and inventory study plans.

2.1. Data Stewardship Categories

The SFAN contributes to the public service mission of the National Park Service and the Inventory and Monitoring Program by providing services to deliver scientific information about park ecosystems. To do this effectively every individual involved with Network projects makes contributions and has stewardship responsibilities in the production, analysis, maintenance, and/or end use of the data. Each of these broad stewardship categories have principal, or 'must-do', responsibilities as well as many potential ancillary tasks. Data stewardship activities and associated responsibilities are outlined in Table 2.1.

Table 2.1 Categories of data stewardship involving all Network personnel.

Stewardship Category	General Responsibilities
Production	Creating data or information from any original or derived source. This includes recording locations, measurements and observations in the field, digitizing source maps, keying in data from a hardcopy source, converting existing data sources, image processing, and preparing and delivering informative products, such as summary tables, maps, charts, and reports.
Analysis	Using data to predict, qualify, and quantify ecosystem elements, structure, and function as part of the effort to understand these components, address monitoring objectives, and inform park and ecosystem management.
Maintenance	Preparing and executing policies, procedures, and activities that keep data and information resources organized, available, useful, compliant, and safe.
End Use	Obtaining and applying available information to develop knowledge that contributes to understanding and managing park resources.

The fundamental role of the park data managers is to facilitate and coordinate data stewardship activities. By communicating and working with all the responsible individuals, creating and maintaining data infrastructure and standards, and understanding program and project requirements, the data managers support and guide people and activities for the purpose of meeting the Network's data management objectives.

2.2. Data Management Roles

Individuals rarely stand alone as successful data producers, analysts, managers, and end users. This may be due, in part, to the current and expected capacity, diversity, and rate of change in information technology. When combined with increasing demand for more detailed, higher quality data and information about natural resources and ecosystem functions, these circumstances normally require a suite of people working with strong coordination to successfully steward data and information resources. Expertise from several areas must come together to ensure that data is collected using appropriate methods, and that resulting data sets, reports, maps, models, and other derived products are well managed, credible, representative, and available for current and future needs.

Because the SFAN I&M projects are primarily managed at the park level, data management at the SFAN is emphasized at this level under the direction of park data managers. The SFAN has therefore foregone the role of a Network data manager employed at many NPS I&M Networks and instead developed the role of a lead data manager. The lead data manager serves as a park data manager with the additional tasks of coordinating data tasks among the park data managers and assisting the Network Coordinator with Network level data tasks.

The current data management structure at the SFAN therefore relies on project managers, the lead data manager, park data managers, and park GIS specialists, but also includes information consumers. With inventory and monitoring projects occurring at a park, multi-park, or network level there is a variety of relationships between these five groups.

Project Managers – Project managers oversee and supervise all phases of an inventory and monitoring project and are the point of contact for that project. Each network assigns two project managers (one lead, one backup) to each project, which may consist of network, park or regional staff. They are responsible for the coordination and supervision of all phases of the project, from raw data collection to data validation and documentation to data analysis and reporting. They are also responsible for complying with the protocol methods and data management plan. They are responsible for the final submission of all products and deliverables. For projects involving contractors and/or cooperators, the project manager is also the Contracting Officers Technical Representative (COTR), and must insure that the contractor complies with the terms of the contract or cooperative agreement. Their active involvement in data management determines the quality and usefulness of the project data and overall success and longevity of the I&M Program.

Park Data Managers – Park Data Managers oversee the development, implementation, and maintenance of data infrastructure and standards for specific parks. Park Data Managers facilitate coordination between projects and protocols to allow for interchange of information wherever possible. Park Data Managers facilitate the long-term storage and maintenance of the data. Park Data Managers work with project managers to design databases, applications, and products, and to facilitate data dissemination.

Lead Data Manager – The Lead Data Manager at the SFAN replaces the Network Data Manager position supported at many of the NPS I&M Networks. The Lead Data Manager

supports the development, implementation, and maintenance of data infrastructure and standards for the Network as a whole. Currently this position also serves as Point of Contact for NPSpecies and NatureBib for all of the parks within the Network. The Lead Data Manager provides a level of coordination to ensure that inventory and monitoring products are gathered and maintained at the network and are accessible for submission to the national I&M office. The Lead Data Manager works with Park Data Managers as part of a network data management team. The GOGA Data Manager currently serves as the Lead Data Manager.

GIS Specialists – The GIS Specialists manage spatial data themes associated with network inventory and monitoring projects, as well as other spatial data related to the full range of park resources. They incorporate spatial data into the GIS. They also maintain standards for geographic data, and are responsible for sharing and disseminating GIS data throughout the network.

Information Consumers – Information consumers may include network coordinators, superintendents, resource managers, science advisors, interpreters, and other information consumers. As the primary targets for the information resulting from the I&M Program, feedback and engaging dialog from these users is essential for an effective program. They provide the input in the types of information needed and suggestions for how it should be presented. They provide supporting continued development and maintenance of our data management capabilities. Their satisfaction with the quality, availability and interpretability of the end products is paramount.

2.3. Data Management Coordination

Collaboration among the National Park Service, other public agencies, universities, and non-governmental organizations is necessary to effectively acquire, apply, and promulgate the scientific knowledge gained in parks. The I&M Program encourages coordination among participants at all levels to help ensure that data collected by NPS staff, cooperators, researchers and others are entered, quality-checked, analyzed, reported, archived, documented, cataloged, and made available for management decision-making, research, and education.

The park data managers work with national I&M Program data management staff and regional resource information management personnel to maintain awareness and involvement in service-wide and regional databases and data management policy and guidance. In addition, the park data managers work locally with Network personnel, park staff, and cooperators to promote and develop workable standards and procedures that result in integration and availability of datasets. Key contacts for the data managers include park GIS specialists and project managers for each monitoring or inventory project. Regular and productive communication among personnel leads to common understanding and better synchronization of Network and park data management activities.

Coordination among the core SFAN data management workgroup is a critical component of meeting SFAN data management objectives. The lead data manager schedules and chairs regular meetings in order to maintain communication within the data management workgroup, provide

updates on service-wide and regional data management policies and guidance, and to facilitate collaboration on SFAN data tasks. The SFAN considers data management to be a team effort and encourages the park data managers and GIS specialists to collaborate on data tasks by reviewing data products, sharing expertise on data management hardware and software, and seeking assistance with data tasks when needed.

Additionally, park and Network staff coordinates on data management through the meetings and work of the Network's Technical Steering Committee and Board of Directors. The development of Network planning materials, including inventory study plans and monitoring protocols, includes involvement and input from park scientists and resource information management staff.

2.4. Shared Responsibilities

2.4.1 Data Stewardship

Keeping track of data from the time of acquisition until it is no longer useful is the shared responsibility of everyone involved with data as a producer, analyst, manager, or end user. This, in essence, is data stewardship. It is a principle of mutual accountability rather than a particular job for one individual. The importance of data stewardship cannot be understated, and the degree of success with which it is used will have direct bearing on the quality and utility of products developed by the Network. However, the project manager oversees the project, supervises technicians conducting field and office work, is responsible for managing the peer review process (with regional and national I&M staff), manages contractors, insures close collaboration with data managers and is therefore ultimately accountable for the project or program's deliverables.

Project managers, data managers, and GIS specialists comprise the central data management team for inventory and monitoring projects. Each is responsible for certain aspects of project data, and all share responsibility for some overlapping tasks. Because of the collaborative nature of project data management, communication among project managers, data managers and GIS specialists is essential to meeting program goals.

2.4.2. Awareness and Accountability

Successful data stewardship requires that people involved in Network activities learn and understand the expectations for continuous data management, and be accountable to perform the duties required to meet these expectations. This is equally important for Network staff, park employees, and contractors or cooperators.

All project participants, including contractors, must receive training, briefings, materials, and additional regular communication about data stewardship from supervisors, project managers, and data managers. The purpose is to promote the appropriate level of understanding about how their efforts relate to park and network management objectives, National Park Service and Department of Interior policies, and other federal government requirements. Other relevant

context and linkages can also be discussed to help establish a sense of ownership and accountability with project staff.

Inventory and Monitoring project managers must have a good understanding of resource information management issues and requirements and be aware of the challenges and limitations of field data collection. This is achieved through training, detailed and regular briefings, and accompanying field crews to perform data collection at reasonable intervals.

2.4.3 Data Stewardship Framework

Appendix A outlines the data stewardship framework adopted by the SFAN. The framework summarizes the programmatic responsibilities, data stewardship responsibilities, and required data awareness levels expected of Network and park level roles. Data stewardship responsibilities are also specified with more detail in individual monitoring protocols and study plans.

3. Data Management Program Overview

3.1. Program Goals

The goal of the SFAN I&M data management program is to ensure the quality, interpretability, security, longevity and availability of natural resource data related to the status and trends of natural resources in the SFAN parks. Natural resource data are the vital building blocks for our evolving ecological understanding about park resources. But a set of data – whether collected the previous year or 20 years ago – must also be accompanied by sufficient context about how and why it was collected to maintain its value beyond the lifetimes of those who collected it. Therefore, a data management program cannot simply attend to the tables, fields, and values that make up a data set. There must also be a process for developing, preserving, and integrating the context that makes it interpretable and valuable.

The term "data" is frequently used in a way that also encompasses other products that are generated alongside the tabular and spatial data that are the primary targets of our data management efforts. These products fall into five general categories: raw data, derived data, documentation, reports, and administrative records (Table 3.1).

Table 3.1. Categories of data products and project deliverables

Category	Examples
Raw data	GPS rover files, raw field forms and notebooks, photographs and sound/video recordings, telemetry or remote-sensed data files, biological voucher specimens
Compiled/derived data	Relational databases, tabular data files, GIS layers, maps, species checklists
Documentation	Data collection protocols, data processing/analysis protocols, record of protocol changes, data dictionary, FGDC metadata, data design documentation, quality assurance report, catalog of specimens/photographs
Reports	Annual progress report, final report (technical or general audience), periodic trend analysis report, publication
Administrative records	Contracts and agreements, study plan, research permit/application, other critical administrative correspondence

To meet I&M program goals – and to ensure adequate context for the primary data products – these various categories of project deliverables all require some level of management to ensure their quality and availability when needed. It is our intent to integrate the manner in which our network creates, manages, and makes available the direct products of our scientific efforts. This requires a holistic view of how natural resource data is generated, processed, finalized and made available. With such an understanding, we can tailor our procedures and infrastructure to meet our objectives without adding unnecessary overhead.

This chapter will describe the various sources of natural resource data, and will present conceptual models for the generalized phases of projects that generate natural resource data. This will provide the framework for describing our data management strategies and infrastructure as they pertain to different project stages.

3.2. Sources of Natural Resource Data

There are many potential sources of important data and information about the condition of natural resources in our parks. The types of work that may generate natural resource data about park resources include:

- Inventories
- Monitoring
- Protocol development pilot studies
- Special focus studies done by internal staff, contractors or cooperators
- External research projects
- Monitoring or research studies done by other agencies on park or adjacent lands
- Resource impact evaluations related to park planning and compliance with regulations
- Resource management and restoration work

Because our program focuses upon long-term monitoring and natural resource inventories, our first priority must be toward the data and information that is derived from these efforts. However, the same standards, procedures, infrastructure and attitudes toward data management can be easily applied to management of natural resource data from other sources. Rather than establishing a separate data management program for the different types of projects, it becomes more an issue of prioritizing and managing the workload and other resources. Naturally, high-profile data sets that provide crucial information to park management should be prioritized for data management regardless of the funding source.

3.3. Data Management Process and Work Flow

3.3.1. Project Work Flow

This section describes the generalized phases of a project in terms of how natural resource data is generated, processed, finalized and made available. Data management activities that relate to the various stages of a project are highlighted. Understanding the life cycle of data throughout a project will help to manage the staffing resources necessary to develop and support quality data.

From the perspective of managing workflow, there are two main types of projects: short-term and long-term. Short-term projects may include individual park research projects, inventories, or pilot work done in preparation for long-term monitoring. Long-term projects will mostly include the implemented monitoring projects central to the I&M program, but may also include multi-year research projects and monitoring performed by other agencies and cooperators. Long-term projects will often require a higher level of documentation, review and infrastructure

development. From a data management standpoint, a primary difference between short- and long-term projects is an increased need to adhere to standards for long-term projects to ensure internal compatibility over time. While the need to follow standards is still present for short-term projects, some times the cost of compliance will outweigh the benefits due to the scope, budget, and level of NPS control of the project. Nevertheless, both short-term and long-term projects share many work flow characteristics, and both generate data products that must be managed and made available.

A project can be divided into five primary stages: planning and approval; design and testing; implementation; product integration; evaluation and closure (Figure 3.1). Each stage is characterized by a set of activities carried out by staff involved in the project.

Primary responsibility for these activities rests with different individuals according to the different phases of a project. These responsibilities are described briefly below under the section for each phase. Additional discussion of the different roles and responsibilities of park and network staff can be found in Chapter 2 of this plan.

3.3.2. Planning and Approval

This first project phase is where many of the preliminary decisions are made regarding project scope and objectives. In addition, funding sources, permits and compliance are all addressed in this phase. Primary responsibility throughout this phase rests with project managers and program administrators. Although there aren't specific data management activities, it is important that data managers remain informed of projects in this phase. This is especially true as timelines for deliverables are finalized. All contracts, agreements and permits should include standard language that describes the formats, specifications, and timelines for project deliverables. Also at this stage, the level of involvement on the part of data management staff should be considered and formalized in the annual work plan.

3.3.3. Design and Testing

Design and testing is where the details are worked out regarding how data will be acquired, processed, analyzed, reported and made available to others. For this reason, an essential collaboration must be developed between the project manager and the data manager. During design and testing, the project manager is responsible for development and testing of project methodology, or modifying existing methods to meet project objectives. During this process, specific procedures related to data acquisition, storage, processing, analysis, and quality control are developed and documented. Ideally, the project manager will seek input and review as these methods are developed and written up, so that integration with existing data management practices, SOPs and guidelines can be accomplished. The data manager can also assist with the process of adapting existing data management SOPs to meet project objectives, and can provide a review of the clarity and completeness of the documentation.

An important part of this collaboration is the development of the data design and data dictionary, where the specific data parameters that will be collected are defined in detail. Devoting adequate attention to this aspect of the project design is possibly the single most important part of assuring

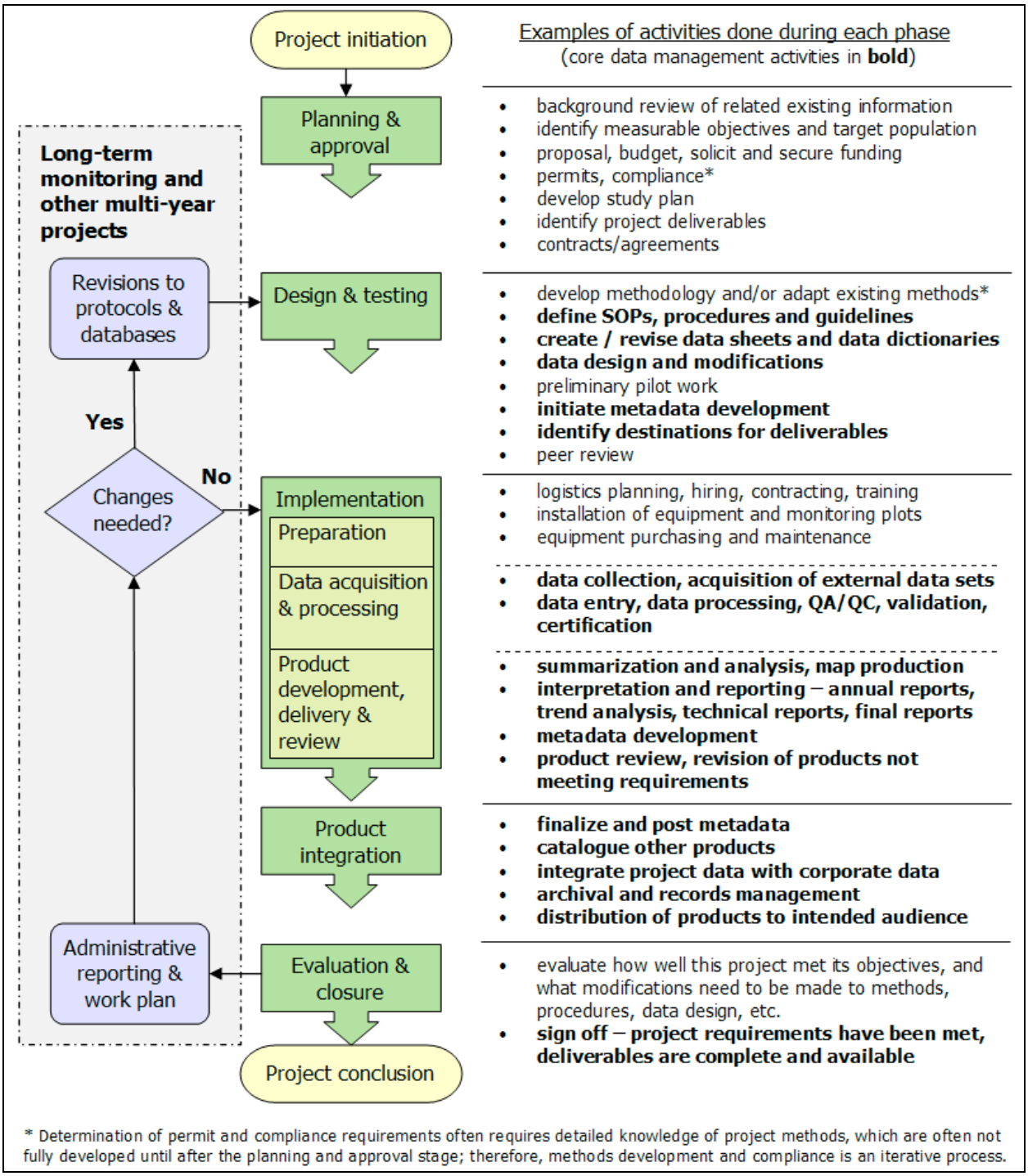


Figure 3.1. Conceptual model of project work flow (Boetsch et al. 2004).

the quality, integrity and usability of the resulting data. By working together on this aspect of project design, the data manager and project manager can take advantage of the different perspectives of the other to the benefit of the quality and interpretability of the resulting data products. In addition, the data manager may be in a position to consult on ways in which minor changes can lead to better opportunities for integration with data from other projects.

All protocol development and data design for I&M monitoring projects requires a peer review process that reaches the regional and national NPS level. This review includes consideration of the data design, quality assurance, and reporting and analysis methods. Similar peer review of other high-profile projects is also advised.

Once the project methods, data design, and data dictionary have been developed and documented, a database can be constructed to meet project requirements. Database development will generally be the responsibility of data managers for I&M funded projects and the project manager for all other projects. The resulting database applications should be thoroughly tested for bugs by entering, editing, summarizing and exporting several dozen test records prior to implementing the database.

At the close of the design and testing phase, project metadata records will be initiated by the project manager. It is also important at this point to identify the destinations for project deliverables and to ensure that they will be produced in a manner that meets program requirements, and decisions should be made regarding integration and permanent storage of these deliverables as they are produced. The project manager will maintain the program specifications for deliverables, and can work with the data manager to identify the formats, special handling requirements, and intended audience for the deliverables, and to refine the timeline for project implementation.

3.3.4. Implementation

During the implementation phase, data are acquired, processed, error-checked and documented. This is also when products such as reports, maps, GIS themes, and other products are developed and delivered. The project manager oversees all aspects of implementation – from logistics planning to data acquisition, report preparation and final delivery. During this phase, data management staff function primarily as facilitators – providing training and support for database applications, GIS, GPS and other data processing applications; facilitation of data summarization, validation and analysis; and assistance with the technical aspects of documentation and product development. The specific roles of data management staff during this phase will depend primarily on the technical capabilities of the project staff. As much as is possible, these roles should be worked out in advance of implementation.

There are three main parts of the implementation phase:

- *Preparation* – This includes all aspects of logistics planning (housing, staging for access to remote sites), hiring and contracting, training, equipment procurement and maintenance, and installation of equipment and monitoring sites. This may also involve preparations for

handling and accessioning specimens, delivery of materials to contract labs for analysis, office or lab space acquisition, etc.

- *Data acquisition and processing* – The exact nature of data acquisition will vary widely from project to project. Many projects will require a significant amount of fieldwork and manual data collection, whereas others may involve only a single trip – or none for projects acquiring data via remote sensing. In some cases, external data sets will be acquired from external sources, either by downloads or other means. Data processing will also vary from project to project, but includes all aspects of data entry and verification for accurate transcription, error-checking and data manipulation, and validation for logical or structural problems with the data. All aspects of data acquisition should be as specified in project protocols and SOPs. Any deviation from these protocols should be carefully documented, such as in the case of equipment failure or adverse weather conditions. Similarly, quality assurance measures should be documented as part of the project metadata.

- *Product development, delivery and review* – The project staff works to develop and finalize the deliverables that were specified in the project study plan (or contract, agreement or permit). These should be developed and delivered according to product specifications. Specifications for such deliverables should be included or referred to in all contracts, agreements, and permits. Products that do not meet program requirements will be returned for revision. All raw and derived data products, metadata, reports and other documentation should be delivered to the data manager. Administrative records should be delivered to appropriate park and network staff as specified for storage.

3.3.5. Product Integration

The data manager is primarily responsible for coordinating product integration. In this phase, data products and other deliverables are integrated into national and network databases, metadata records are finalized and posted in clearinghouses, and products are distributed or otherwise made available to its intended audience. This is also when items that belong in collections or document archives are accessioned and catalogued.

Product integration includes creating records for reports and other project documents in NatureBib, posting imaged documents to the appropriate repository, posting metadata records that have been completed and submitted by project managers, and updating NPSpecies to reflect any new species occurrence information derived from the project. This will allow the information from the project to be searchable and available to others via service-wide search engines.

Another aspect of integration is merging data from a working database to a corporate database maintained on the network server. This occurs only after the annual working dataset has been certified for quality by the project manager. Certain projects may also have additional integration needs, such as when working jointly with other agencies for a common database.

3.3.6. Evaluation and closure

For long-term monitoring and other cyclic projects, this phase occurs at the end of each field season, and leads to an annual review of the project. For non-cyclic projects, this phase represents the completion of the project. After products are catalogued and made available, program administrators, project managers, and data managers should work together to assess how well the project met its objectives, and to determine what might be done to improve various aspects of the methodology, implementation, and formats of the resulting information. For I&M protocols, careful documentation of all changes is a requirement. Changes to methods, SOPs and other procedures are tracked in separate format associated with each document. Major revisions may require additional peer review. Upon project closure, records are updated to reflect the final status of the project and its associated deliverables in the network and park archives.

4. Data Management Infrastructure and Architecture

The SFAN computer resource infrastructure is composed of computers and servers that are functionally or directly linked through computer networking services. This infrastructure represents the foundation upon which our network information system is built. Systems architecture signifies the applications, database systems, repositories, and software tools that make up the framework of our data management enterprise.

The SFAN I&M program relies heavily on park, regional, and national IT personnel and resources to maintain the computer resource infrastructure. This includes but is not limited to hardware replacement, software installation and support, security updates, virus-protection, telecommunications networking, and backups of servers. Therefore communication with park and regional IT specialists is essential to ensure adequate resources and service continuity for our systems architecture. This chapter details a snapshot of our current computer resources and describes our computer infrastructure in general terms, focusing more specifically on the systems architecture that is central to data management.

4.1. Computer Resources Infrastructure

An important element of a data management program is a reliable, secure network of computers and servers. The SFAN digital infrastructure has three main components: servers maintained at the national level, park-based local area networks (LAN), and an network directory nested within the GOGA LAN. This infrastructure is maintained by park, regional, and national IT specialists, who administer all aspects of system security and backups.

These components each host different parts of our natural resource information system.

National Servers

- Master applications – integrated client-server versions of NatureBib, NPSpecies, NR-GIS Metadata Database
- Centralized repositories – NR-GIS Data Store, Protocol Clearinghouse
- Public access sites – portals to NatureBib, NPSpecies, NPSFocus, and websites for monitoring networks

Park LANs

- Local applications – desktop versions of national applications such as NPSpecies and Dataset Catalog
- Working files – working databases, draft geospatial themes, drafts of reports, administrative records
- Park digital archives – base spatial data, finalized datasets, and finished versions of park project deliverables
- Park GIS files – base spatial data, imagery, and project-specific themes

Network Directory

- Master project databases – compiled data sets for monitoring projects and other multi-year efforts that have been certified for data quality
- Network digital archives – network repository for finished versions of project deliverables for I&M projects (e.g., reports, methods documentation, data files, metadata, etc.)

4.1.1. SFAN Computer Hardware

In order to create and maintain spatial and tabular data for the I&M program, park data managers and Network staff require a robust set of computer hardware components and devices. Table 4.1 lists the computer hardware employed by the SFAN as of October 2004.

Table 4.1. SFAN computer hardware resources.

COMPUTER HARDWARE	QUANTITY
Cisco 3640 router	1
Cisco 2924XL Switch	1
Transition Network fiber transceiver	1
Dell PowerEdge 4600 Server	1
Dell Optiplex GX200	6
Dell Optiplex GX270	2
Dell Pentium III Laptop	2
Dell C600 Laptop	3
Dell C840 Laptop	2
Trimble GeoIII	1
Trimble GeoXT	2

4.1.2. SFAN Computer Software

In addition to computer hardware, an array of computer software programs is needed for park data managers and Network staff to meet the data management goals of the I&M program. Table 4.2 lists the computer software employed by the SFAN as of October 2004.

Table 4.2. SFAN software and computer resources.

APPLICATION/FUNCTION	SOFTWARE PACKAGE/VERSION	QUANTITY
Web Development	Macromedia Dreamweaver MX 2004	2
Web Development	Cold Fusion MX	1
Operating System	Microsoft Windows	15
Database	Microsoft Access XP	15
Word Processing	Microsoft Word XP	15
Spreadsheet	Microsoft Excel XP	15
Presentation	Microsoft PowerPoint XP	15
	Adobe Acrobat	2
GPS	Trimble Pathfinder Office	3
GIS Desktop	ArcView 3.3, ArcGIS 8.3, ArcInfo	4

APPLICATION/FUNCTION	SOFTWARE PACKAGE/VERSION	QUANTITY
Statistics	PCOrd	1
	StatView	1
Field Data Entry	Pendragon	1
Image Processing	ERDAS Imagine, Orthobase	1
Image Compression	Mr SID	1
Virus Protection	Symantec Anti-Virus Corporate Version	1

4.1.3. Digital File Structures

Each of the SFAN Park LANs accommodates hierarchical directory structures for storing digital files. Because the majority of the SFAN I&M projects are managed at the park level, working files, including protocols, databases, GIS coverages, and reports are maintained on the Park LANs. Digital files are only sent to the Network level under annual reporting guidelines or as products are finalized. As such, it is necessary to document file folder structures for each of the Park LANs as well as the SFAN directory. File folder structures are presented in Appendix B.

4.1.4. Server Back-up Specifications and Schedules

The risk of data loss can come from a variety of sources, including catastrophic events (e.g., fire, flood), user error, hardware failure, software failure or corruption, and security breaches or vandalism. Performing regular backups of data and arranging for off-site storage of backup sets are the most important safeguards against data loss.

With working and archived I&M project files stored on Park LANs and on the Network directory, it is important to outline all of the relevant server back-up procedures within the SFAN.

Golden Gate National Recreation Area

The Dell PE 4600 server in the Marin Headlands is a critical component of the SFAN computer infrastructure. The server houses all pertinent digital files for resource management personnel stationed in the northern sector of GOGA at Building 1061 at Fort Cronkite, including the park's Wildlife Biologist, Aquatic Ecologist, and Hydrologist. Because the GOGA component of the I&M program is managed from this station, all of GOGA's digital I&M files are stored on the Headlands server.

A separate directory on the Headlands server houses the Network directory. This directory stores all of the SFAN central files, including the Network digital archives.

The server has several redundancy built to ensure data are kept safe. To start off, the operating system (OS) Windows 2000 server is installed on a pair of hard drives that are 16GB each in size and configured as level-1 RAID (mirroring). The data are stored on a level-5 RAID with data capacity of 200GB. Additionally, the server is configured with 4 redundant power supply

connecting to an APC Smart Ups 2200XL uninterruptible power supply with the capability to shut the server down during power outages.

A SDLT 110/220 GB tape drive is also connected to back up data. The backup scheme being implemented is full nightly backup on every Tuesday and Friday. The remaining days of the week including Saturday and Sunday are incremental and are appended to the same tape that is used for full backup. The last tape of the month is taken offsite to further protect the data. Only GOGA IT team members will be responsible for managing the tape backup jobs and tapes.

Physically the server itself is located at Marin Headland in Building 1061. The server is locked and only GOGA IT team members have access.

Data security to the I&M folder is configured with only I&M members and GOGA IT having access. All users are denied any access. All requests to add new members require Supervisor approval or someone designated by the Supervisor managing the folder to send an email to GOGA IT. This email will be kept as a record of all changes.

Pinnacles National Monument

The I&M data at Pinnacles National Monument is stored and backed-up on a series of four Snap Appliance, Snap Server 1100's. The Snap Servers were first introduced to the Research and Resource Management office in approximately 2001. These servers have Auto-sensing 10/100Base-TX, single RJ-45 network connectors and supports DHCP, BOOTP and RARP for automatic assignment of IP addresses.

The (RRM-Data) snap server acts as the working directory and primary data storage directory for all Research and Resource Management and Inventory and Monitoring documents and projects excluding images. In the event that the (RRM-Data) snap server crashes or fails, a backup copy of these data are located on the (RRM-Backup) snap server.

I&M image data is stored on the (RRM-Images) server. This server acts as the working directory and primary data storage for I&M and RRM image data. These data have a backup located on the (RRM-Backup2) snapserver. Data saved on either the RRM-Backup or RRM-Backup2 snap servers can easily be restored with windows XP or with Powerquest Datakeeper Snap Appliance software. Currently the backup scheme being implemented is a full nightly backup Monday thru Friday.

Physically the Snap Servers are located in the Research and Resource Management office at Pinnacles National Monument. Only personnel that have access to the Research and Resource Management building can access the servers. Due to facility limitations, the Snap Servers can be accessed by anyone in this building.

Point Reyes National Seashore

A Dell 715N Network Attached Storage (NAS) server (160GB) holds the I&M data and information specific to PORE along with non I&M files. All GIS data at PORE is stored on a Dell 725N PowerValult NAS server (320GB). These data storage units are configured as RAID5

as a first level of protection in the event of a drive failure. Both units are connected to un-interruptible power supplies capable of implementing a software driven shut down in the event of an extended power failure (greater than 10 minutes). Both these units reside in the LAN room in the administration building at Park Headquarters.

The LAN domain controller, a Dell PowerEdge 4400, resides in the LAN room in the Resource Management Building and runs the park's automated network backup system using the BrightStor ARCserve Backup r11.1 (Computer Associates http://ca.com) application. A Sony AIT II tape drive, attached to the server is capable writing 50GB to a single tape. ARCserve Backup is configured with a Grandfather, Father, Son (GFS) backup scheme. Full backups occur every Friday, with the last Friday of the month considered the monthly backup. Incremental backups occur Monday through Thursday, weekends there is no backup. Weekly and monthly backup tapes are stored offsite, in a fire resistant safe, with weekly tapes being recycled after the next monthly is created. Monthly tapes are retained for 4 months before being recycled.

PORE is currently in the process of switching to a Granite FIREVue FireWire Hot-swappable drive 8 bay jukebox back-up device. This hardware will replace the Sony AIT II tape drive with a jukebox containing a series of 8 180GB hot-swappable IDE hard drives. PORE has an additional 8 180GB drives to fully implement the GFS backup scheme. ARCserve Backup will remain the software package that implements PORE's network backup. We expect to have this new system running by Febuary 2005.

4.2. National Information Management Systems

The need for effective natural resource information management cuts across NPS divisional boundaries and management strategies must be defined at the highest level possible. In this context, integrated inventory and monitoring of natural resources is multidisciplinary and requires national-level, programmatic data and information management strategies for success.

The basic strategy of natural resource and therefore inventory and monitoring information management is to provide integrated natural resource databases and information systems that enhance NPS managers' and staff's access and use of timely and valid data and information for management decisions, resource protection, and interpretation. Inventory and monitoring information needs are broadly separated into two categories:

- *Detailed data and information needed for onsite resource management and protection*. The information used to guide natural resource management decisions must be specific to inform and be useful to management staff at parks and central offices.
- *Summary information needed to describe the resources and their condition*. This kind of information usually needs to be aggregated across the National Park Service for use by NPS and DOI managers and central office personnel to answer requests from Congress and for budget, program, and project planning.

The NPS Natural Resource Program Center (NRPC) and the I&M Program actively develop and implement a national-level, program-wide information management framework. NRPC and

I&M staff integrate desktop database applications with internet-based databases to serve both local and national-level data and information requirements. Centralized data archiving and distribution capabilities at the NRPC provide for long term data security and storage. NRPC sponsors training courses on data management, I&M techniques, and remote sensing to assist I&M data managers with developing and effectively utilizing natural resource information.

4.2.1. National-level application architecture

To achieve an integrated information management system, three of the national-level data management applications (NatureBib, NPSpecies, and NR-GIS Metadata Database) utilize a distributed application architecture with both desktop and internet-accessible (master) components (Figure 4.2).

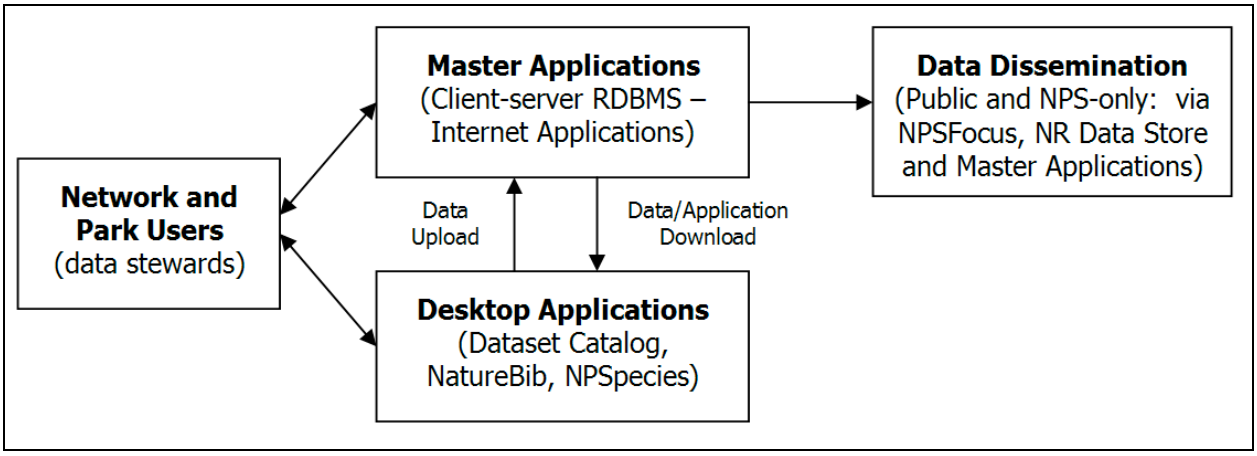


Figure 4.2. Model of the national-level application architecture.

NatureBib

NatureBib is the master database for bibliographic references that merges a number of previously separate databases such as Whitetail Deer Management Bibliography (DeerBib), Geologic Resource Bibliography (GRBib), and others. It also contains citation data from independent databases like NPSpecies and the Dataset Catalog and NR-GIS Metadata Database. It currently focuses on natural resource references, but may eventually be linked to references on cultural

resources and other park operations. As with NPSpecies and NR-GIS Metadata Database, it is possible to download data from the master web version into the MS Access desktop version that can be used locally on computers with limited internet connectivity (*http://www.nature.nps.gov/nrbib*).

NPSpecies

NPSpecies is the master species database for the NPS. The database lists the species that occur in or near each park, and the physical or written evidence for the occurrence of the species (e.g., references, vouchers, and observations). Taxonomy and nomenclature are based on ITIS, the interagency Integrated Taxonomic Information System. The master version of NPSpecies for each park or network can be downloaded from the master website into an MS Access version of NPSpecies. The internet-based version is the master database, which can be accessed via password-protected logins administered by park, network and regional data stewards assigned for each park and network. The master database requires that species lists are certified by networks before any data will be available to the public. NPSpecies is linked to NatureBib for bibliographic references that provide written evidence of a species' occurrence in a park and will be linked to NR-GIS Metadata Database to document biological inventory products. The MS Access application and additional details can be found at the NPSpecies website (*http://science.nature.nps.gov/im/apps/npspp/index.htm*).

Dataset Catalog and NR-GIS Metadata Database

Dataset Catalog is a desktop metadata database application developed by the I&M Program to provide a tool that parks, networks, and cooperators can use to inventory and manage data set holdings. Although not designed as a comprehensive metadata tool, the Dataset Catalog is used for cataloging abbreviated metadata about a variety of digital and non-digital natural resource data sets. The Dataset Catalog helps parks and networks begin to meet Executive Order 12906 mandating federal agencies to document all data collected after January 1995. It provides brief metadata and a comprehensive list about all resource data sets for use in data management, project planning, and more stringent metadata activities. As with other service-wide applications, the master metadata database (NR-GIS Metadata Database) is available through a website and will be linked to NPSpecies (the NPS species database) and NatureBib (the bibliographic database). It will be possible to download a version in MS Access format from the master website (*Dataset Catalog: http://science.nature.nps.gov/im/apps/datacat/index.htm and NR-GIS Metadata Database: http://science.nature.nps.gov/nrdata*).

Other national-level I&M information management and GIS applications include:

NPSTORET

STORET is an interagency water quality database developed and supported by the Environmental Protection Agency's (EPA) to house local, state, and federal water quality data collected in support of managing the nation's water resources under the Clean Water Act. STORET is used by the NPS as a repository of physical, chemical, biological, and other monitoring data collected in and around national park units by park staff, contractors, and cooperators. The NPS operates its own service-wide copy of STORET and makes periodic uploads to the EPA STORET National Data Warehouse so that data collected by and for parks will be accessible to the public. NPS Director's Order 77 indicates that the NPS should archive

water quality data in STORET, and the NPS Water Resources Division (WRD) requires that any data collected as part of a funded WRD project get archived in STORET. NPSTORET (also known as Water Quality Database Templates) the NPS master database designed to facilitate park-level standardized reporting for STORET. The database is still in development, but metadata, protocols, data dictionaries, and reporting capabilities are available through a front-end form. Upon implementation, network staff and cooperators will be able to use the MS Access version of NPSTORET either as a direct database for data entry and management, or as a means of submitting data for upload to STORET by WRD staff. The MS Access application and additional details can be found at: *http://www.nature.nps.gov/water/infodata.htm*. Additional information on STORET can be found at: *http://www.epa.gov/storet*.

Natural Resource Database Template

The Natural Resource Database Template (NRDT) is a flexible, relational database in MS Access for storing inventory and monitoring data (including raw data collected during field studies). This relational database can be used as a standalone database or in conjunction with the GIS software (e.g., ArcView or ArcGIS) to enter, store, retrieve, and otherwise manage natural resource information. The template has a core database structure that can be modified and extended by different parks and networks depending on the components of their inventory and monitoring program and the specific sampling protocols they use. Natural Resource Database Template is a key component of the I&M program's standardized monitoring protocols. These monitoring protocols include separate modules detailing different aspects of monitoring project implementation, from sampling design to data analysis and reporting, and include data management components that describe database table structure, data entry forms and quality checking routines. Approved monitoring protocols, including the databases that are based on the Database Template, are made available through a web-based protocol clearinghouse (see below). A description of the Database Template application, a data dictionary, and example implementations are located on the NR Database Template website (*http://science.nature.nps.gov/im/apps/template/index.htm*).

Natural Resource Monitoring Protocol Clearinghouse

The Natural Resource Monitoring Protocol Clearinghouse (i.e., Protocol Database) is a web-based clearinghouse of sampling protocols used in national parks to monitor the condition of selected natural resources. The database provides a summary of, and in many cases allows the user to download a digital copy of, sampling protocols that have been developed by the prototype monitoring parks or other well-established protocols used in national parks. The Protocol Database also makes it possible to download database components (e.g., tables, queries, data entry forms) in MS Access that are consistent with the Natural Resource Database Template that have been developed for a particular protocol. See the Protocol Database website for available protocols (*http://science.nature.nps.gov/im/monitor/protocoldb.cfm*).

NR-GIS Data Store

The NR-GIS Data Store is a key component of the data dissemination strategy employed by the I&M Program. The NR-GIS Data Store is a graphical search interface that links dataset metadata to a searchable data server on which datasets are organized by NPS units, offices and programs. The interface allows customized public or protected searches of natural resource datasets, inventory products and GIS data produced by the I&M and Natural Resource GIS

Programs. Each park or network is able to post and curate its data on the server. The NR-GIS Data Store will be integrated with the master NR-GIS Metadata Database application to streamline programmatic data documentation and dissemination processes. The simple browse function of this server can be accessed at: *http://nrdata.nps.gov/*.
See the NR-GIS Data Store website for further information (*http://science.nature.nps.gov/nrdata*).

4.3. Network Systems Architecture

Rather than developing a single, integrated database system, our data design relies upon modular, standalone project databases that share design standards and links to centralized data tables. Individual project databases are developed, maintained, and archived separately. There are numerous advantages to this strategy:

- Data sets are modular, allowing greater flexibility in accommodating the needs of each project area. Individual project databases and protocols can be developed at different rates without a significant cost to data integration. In addition, one project database can be modified without affecting the functionality of other project databases.
- By working up from modular data sets, we avoid a large initial investment in a centralized database and the concomitant difficulties of integrating among project areas with very different – and often unforeseen – structural requirements. The payoff for an initial investment into an integrated database may not be realized down the road by greater efficiency for interdisciplinary use.

4.3.1. Database Design Standards

Project database standards are necessary for ensuring compatibility among data sets, which is vital given the often unpredictable ways in which data sets will be aggregated and summarized. When well thought out, standards also help to encourage sound database design and facilitate interpretability of data sets.

Relational database guidelines adopted by the SFAN I&M Program are drawn from the Natural Resource Database Template. The database system is composed of multiple, separate project databases designed in MS Access with a similar database structure. Each project database functions independently, but common database elements allow data to be combined from multiple databases for comprehensive analysis and reporting purposes. The SFAN database design standards are outlined in more detail in Appendix C.

Databases that are developed for park and network projects will all contain the following main components:

- *Common lookup tables* – Links to entire tables that reside in a centralized database, rather than storing redundant information in each database. These tables typically contain information that is not project-specific (e.g., lists of parks, personnel, and species).

- *Core tables and fields based on network and national templates* – These tables and fields are used to manage the information describing the "who, where and when" of project data. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. These core tables contain critical data fields that are standardized with regard to data types, field names, and domain ranges.
- *Project-specific fields and tables* – The remainder of database objects can be considered project-specific, although there will typically be a large amount of overlap among projects. As much as is possible, efforts will be made to develop these project-specific objects to be compatible with those maintained by other networks and cooperators managing similar data sets – especially if integration with other data sets is important for meeting project objectives.

It is important to emphasize the need to store common lookup tables in a centralized database. Because lists of contacts, projects, parks, species are often complex and dynamic, it is a good strategy to centralize this information so that users have access to the most updated versions in a single, known place. Centralizing also avoids redundancy and versioning issues among multiple copies. Centralized information is maintained in database tables that can be linked or referred to from several distinct project databases. Network applications – for project tracking, administrative reporting, or budget management – can also link to the same tables so that all users in the network have instantaneous access to edits made by other users.

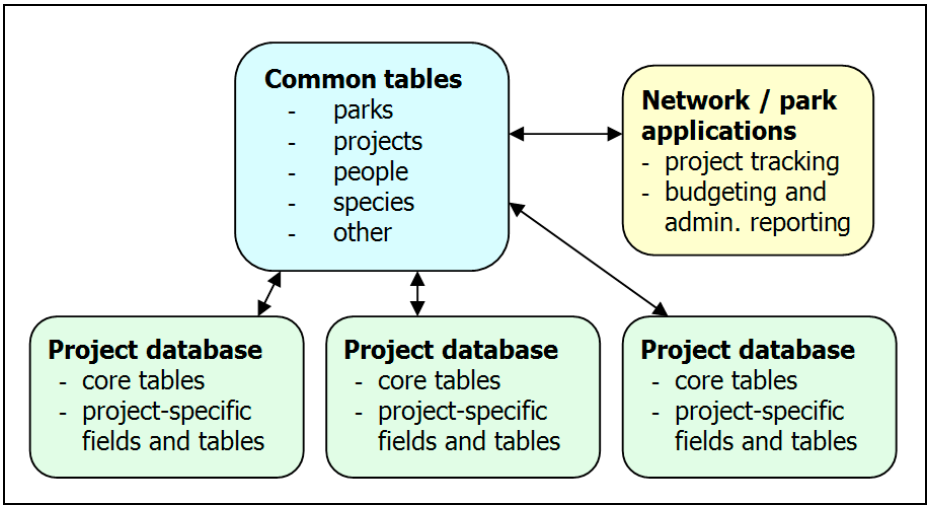


Figure 4.3. Common lookup tables and satellite databases (Boetsch et al. 2004).

Separating these tables by functional groupings is done primarily to reduce conflicts and performance losses associated with multiple users in MS Access. Databases associated with individual projects each access the common tables via links established in each project back-end data file.

4.3.2. Levels of data standards

The three types of database objects (common lookup tables, core tables, and project-specific tables) also correspond to three putative levels of data standards. Because common lookup tables are stored in one place and are referred to by multiple databases, they represent the highest level of data standard because they are implemented identically among data sets. The second

level of standards is implied by the core template fields and tables, which are standardized where possible, but project-specific objectives and needs could lead to varied implementations among projects. The third level of standards is applied most flexibly to accommodate the range of needs and possibilities for each project, yet always with compatibility and integrity in mind. The following figure presents the resulting variation in implementation of these differing levels as a "bull's eye", with the common lookup tables providing the most consistent implementation and hence the smallest range of variation.

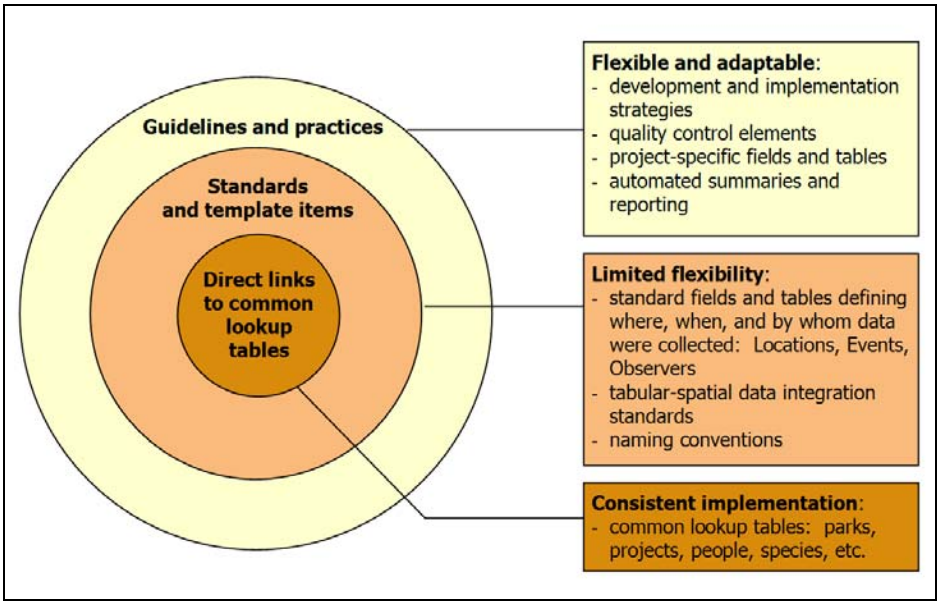


Figure 4.8. Different levels of data standards and their corresponding degree of implementation variability (Boetsch et al. 2004).

5. Data Acquisition and Processing

The National Park Service's I&M Program is responsible for acquiring the necessary information required by park managers to properly manage and maintain the natural resources of their park. To successfully accomplish this task, information from multiple sources is collected by the SFAN Inventory and Monitoring Program and processed to ensure that it meets the data standards established by the SFAN. This chapter describes the initial steps in a progression of steps necessary to strengthen the scientific foundation of the I&M program and provide the park managers with high quality information to manage park resources.

There are three general classifications for the types of data handled by the I&M Program:

1. **Programmatic Data** – any data produced from projects that are initiated (funded) by the I&M Program or projects that in some way involve the I&M Program.

2. **Non-programmatic Data**

    - **Non-programmatic NPS Data** – any data produced by the NPS that did not involved the I&M Program.

    - **Non-programmatic Non-NPS Data** – any data produced by agencies or institutions other than the National Parks Service.

The following sections outline in more detail the manner in which the different data types described above are acquired and processed.

5.1. Programmatic Data Sources

The collection of programmatic data under the purview of the SFAN I&M Program is connected to either natural resources inventories or to vital signs monitoring projects.

Natural Resources Inventories

- Resource inventories were designed to identify the primary resources of each park and as such they represent an important phase in the management of park resources. Appendix D lists the 24 basic inventories conducted by the Inventory and Monitoring Program.

Vital Signs Monitoring Projects

- Vital signs are measurable, early warning signals that indicate changes that could impair the long-term health of natural systems. Early detection of potential problems allows park managers to take steps to restore ecological health of park resources before serious damage can happen.

- Each Network, working in conjunction with park resource managers, was given the task of prioritizing the natural resources vital signs that were most important to their parks and to the network as a whole
- The SFAN I&M Program and the Network parks identified and ranked 63 vital signs, of which the first 18 will receive funding and/or support from the I&M program. These 18 vital sign monitoring programs are identified in Appendix E.

Data for these projects are typically collected by I&M personnel, park staff or by cooperators/contractors and include both data discovery and field study components.

5.1.1. Data Discovery

Data discovery, or data mining, is the process of searching for existing data/information that is relevant to the mission I&M Program and to the natural resources of the parks in the Network.

A large percentage of data discovery occurs at the onset of new projects or during the development of new protocols. The process involves reviewing many different sources for varying types of information. Many of the following data sources are accessible via the Internet but some require visiting local research or academic institutions, museums or local parks to conduct the searches.

Bibliographic/Literature

- National NPS Databases (e.g. NatureBib)
- Online literature databases (e.g. First Search or Biosis)
- Local document library
- Library catalogs (e.g. academic or research institutions)
- Park archives

Geographic Data

- Regional GIS Specialists
- Park GIS Specialists
- Geographic data clearinghouses (e.g. USGS, NPS or FWS)
- Dataset Catalog

Biological/Natural Resources Data

- NPSpecies
- Voucher collections
- Network Parks
- Dataset Catalog

Much of the information collected is likely to be *legacy data*, or data that is not current in nature. If legacy data is collected in a digital format the information should be maintained in up-to-date formats that are compatible with the current software standards. In the past, hard copy materials

were maintained as such but in the future, the SFAN intends to scan hardcopy references and materials, saving them as .PDF files, in order to create a digital library. All hard copy reports are also entered into the SFAN bibliographic database, which is based on the desktop version of NatureBib.

5.1.2. Field Studies

A field study is considered any project requiring the collection of new (original) data based on a scientific protocol or study plan as opposed to the collection of existing information which will be discussed later in this chapter. The most common examples of field studies are biological inventories and long-term monitoring initiatives. All such projects involving the I&M Program are supported by the park data managers. Data managers are responsible for ensuring that data collection, data entry, verification, storage, and archiving for any field project are consistent with the SFAN I&M Program standards. In addition to general standard operating procedures (SOPs) that define network-wide requirements, protocol specific SOPs are developed detailing procedures and/or methodologies specific to each protocol. Data managers will work closely with the principal investigators and Network staff to develop these guidelines and methodologies. This may range from detailing the proper usage of data entry forms to outlining calibration procedures for automated data loggers.

5.1.3. Changes to Data Collection Procedures/Protocols

Changes to data collecting procedures are discouraged unless there are important reasons for altering methodologies. Ideally, all problems should be identified during the design and testing stages of the project and changes implemented prior to the collection of any field data. Protocols should attempt to identify any foreseeable issues that might occur as well as contingencies to address them. Inevitably, unforeseeable problems could occur and procedures/protocols would require alteration after data collection has begun. Significant changes to the protocols must be approved by the principal investigator, who will also determine if additional peer review is required before accepting any protocol changes.

Altering data collection procedures or protocols can also occur as a result of a comprehensive review which all monitoring protocols should undergo every five years. During the review data are analyzed to determine if the current protocol accomplished its goal. If it is concluded that the protocol in its present form has not achieved the desired results changes could be recommended. Once again, all changes must be approved by the principal investigator.

5.2. Non-Program Data Sources

Projects implemented by or data products produced by entities other than the Inventory and Monitoring Program are considered *non-programmatic data sources*. Two distinct kinds of non-programmatic data were defined in the introduction of this chapter: (1) *NPS* and (2) *non-NPS programmatic data*. The following section will further describe the data that fall into these classes and the manner in which the SFAN manages such data.

5.2.1. Non-Programmatic Data – NPS Data

A large percentage of data collected in Network parks is collected by the park personnel involved in projects initiated at the individual park level. Regional funding sources such as Natural Resources Protection and Preservation (NRPP) funds are available to conduct park-level or multi-park projects. The data produced by such projects are important and very relevant to the mission of the I&M Program. Network parks often conduct their own park-based inventory projects, the data from which can be used to supplement Network-level inventories conducted by the I&M Program. Parks also engage in monitoring projects which produce information that is very valuable when developing Network-level monitoring protocols.

NPS regional programs also provide a good resource for natural resources information. Programs such as Exotic Plant Management Teams (EPMT) collect and maintain data regarding the presence of exotic species in parks and the methods used to treat these species. This information is stored in the national APCAM (Alien Plant Control and Monitoring Database) which is maintained by the EPMT data manager.

It is important that park, regional and network staff work closely together to ensure that information is maintained in a manner that promotes data sharing. Accordingly, the SFAN data managers will:

- Work closely with park and regional personnel to ensure that high quality data available.
- Provide training to park staff interested in learning to use NPSpecies and NatureBib.
- Develop and/or provide training in developing databases based on the NRDT that meet the needs of park resource managers.

5.2.2. Non-Programmatic Data - External Sources

Data produced from projects that are not conducted in conjunction with the National Park Service or the I&M Program are considered to be *non-programmatic external data.* As was the case in the previous section, external data sources often provide relevant information important to the mission of the I&M Program. It should be noted that such sources need not be directly connected to Network parks but may instead pertain to methodologies or protocols that could assist Network personnel with the development of a more productive program. A large percentage of external data products are in the form of published reports or papers; however, unpublished information such as organized volunteer surveys (e.g. the Christmas Bird Count) or information relating to ambient conditions (e.g. weather/atmospheric data) provides an indication of long-term regional trends. Relevant data such as these should be obtained from the agency or individual responsible for collecting and maintaining the data. The park data managers are responsible for ensuring that the information is properly documented and stored.

6. Quality Assurance/Quality Control

6.1. Importance of Quality Assurance/Quality Control (QA/QC)

The view that the ecological data and related information resulting from SFAN I&M efforts are a valuable resource worthy of preservation is justified only if those data may be used with confidence. Analyses performed to detect trends or patterns in ecosystem processes and the condition of natural resources require data of documented quality that are free from error and bias. Data of poor quality can result in loss of sensitivity to subtle changes and incorrect interpretations and conclusions, and the potential for problems with data quality increases dramatically with the size and complexity of the data set (Chapal & Edwards 1994). Therefore, the most important goal of the SFAN I&M data management program is to ensure that I&M projects produce data of the highest possible quality upon which to base park resource management decisions and that the long-term quality and integrity of the data are maintained.

Documented methods that ensure data quality are critical for safeguarding data integrity. Established protocols for the identification and reduction of error at all stages in the data lifecycle, including project planning, data collection, data entry, verification and validation, processing, and archiving, should be incorporated into the data management infrastructure and institutionalized.

While a data set containing zero errors is the ideal, perfection is rarely achieved, and the cost of attaining 100% accuracy may outweigh the benefit. Two parameters should be considered in setting a data quality goal: 1) the percent of entries that are incorrect (frequency of errors) and 2) the magnitude of the error (criticality of errors). For example, a two-digit numeric entry off by one decimal place is a significant error. On the other hand a six-digit numeric entry with the sixth digit off by one is an insignificant error, having an accuracy of up to 99.999 percent. In another case, one incorrect digit in a six-digit species number indicates a completely different species and is quite significant. Error significance, therefore, is dependent on the type of data. The overall data quality goal should be a reasonable and attainable level of quality based on the intended use of the data and the potential consequences of making a wrong decision.

6.2. NPS Mandate for Quality

Although the functional lifetime of hardware and software is decreasing rapidly, properly archived data are forever. Producers and users must know and document the quality of their data. This is especially important for sharing data and is the intent of several government directives. NPS Director's Order #11B: Ensuring Quality of Information Disseminated by the National Park Service was issued in 2002 to comply with these directives to ensure and maximize the quality of information disseminated by Federal agencies. The order defines *'quality'* as an encompassing term comprising objectivity, utility, and integrity; therefore '*quality*' generally refers to all three of these elements. *'Objectivity'* includes two distinct elements: 1) presentation, whether disseminated information is being presented in an accurate, clear, complete, and unbiased manner within a proper context and 2) substance, a focus on

ensuring accurate, usable, and reliable information. *'Utility'* refers to the usefulness of the information to its intended users, from the perspectives of both the office and the public. *'Integrity'* refers to the security of information, e.g., protection from unauthorized access or revision to ensure that the information is not compromised through corruption or falsification. The order further specifies that information will be developed only from reliable data sources and that it will be accurate, timely, and representative of the most current information available. These standards apply not only to NPS-generated information, but also to information provided by other parties to the NPS if the NPS disseminates or relies upon this information.

High quality data and information are not only mandated by directives and orders, they are vital to the credibility and success of the I&M program. According to Abby Miller (2001) of the Natural Resource Stewardship and Science Division, "data need to meet national-level quality standards and need to be accessible to be used for wise and defensible decision-making at all levels. Data need to be able to be shared and aggregated with data from other parks and from adjacent lands to support landscape-level and national planning and decision-making."

6.3. Definition of QA/QC

Quality assurance procedures plan for quality in all stages of the data development process, while quality control procedures monitor or evaluate the resulting data products. Palmer (2003) defines quality assurance as "an integrated system of management activities involving planning, implementation, documentation, assessment, reporting, and quality improvement to ensure that a process, item, or service is of the type and quality needed and expected by the consumer." He defines quality control as "the overall system of technical activities that measures the attributes and performance of a process, item, or service against defined standards to verify that they meet the stated requirements established by the customer."

QA/QC mechanisms are designed to prevent data contamination, which occurs when a process or event other than the one of interest affects the value of a variable and introduces two fundamental types of errors into a data set. Errors of commission include those caused by data entry and transcription errors or malfunctioning equipment. They are common, fairly easy to identify, and can be effectively reduced upfront with appropriate QA mechanisms built into the data acquisition process, as well as QC procedures applied after the data have been acquired. Errors of omission often include insufficient documentation of legitimate data values, which could affect the interpretation of those values. These errors may be harder to detect and correct, but many of these errors should be revealed by rigorous QC procedures.

QA/QC procedures applied to ecological data include four activities ranging from simple to sophisticated, inexpensive to costly: 1) defining and enforcing standards for electronic formats, locally defined codes, measurement units, and metadata, 2) checking for unusual or unreasonable patterns in data, 3) checking for comparability of values between data sets, and 4) assessing overall data quality. Much QA/QC work is related to the first activity, which begins with data design and continues through acquisition, entry, metadata development, and archiving. The progression from raw data to verified data to validated data implies increasing confidence in the quality of the data through time.

Table 6.1. Quality assurance and quality control procedures that are associated with data design, data acquisition, metadata development and data archival phases in a comprehensive data management system (Brunt 2000).

Quality assurance and quality control (QA/QC)	Design	Acquisition	Metadata	Archive
Check that data sheets represent experimental design	X			
Check that measurement units are defined on the data sheet	X			
Check that attribute names meet project standards	X			
Check that date, site, and coded values meet project standards	X			
Check that attribute names and descriptions are provided	X			
Check that data are complete		X		
Check that data entry procedures were followed		X		
Check that data include time, location, and collector(s)		X	X	X
Check that measurement data are within the specified range		X		
Check that data values or codes are represented correctly		X		
Check that data are formatted correctly for further use		X	X	X
Check that data table attribute names are reasonable		X	X	X
Check that data table design reflects experimental design		X	X	X
Check that values for each attribute are represented one way		X	X	X
Check that errors and corrections are recorded		X	X	X
Check that metadata are present			X	X
Check metadata for content (accuracy and completeness)			X	X
Check that data dictionary is present and accurate			X	X
Check that measurement units are consistent		X	X	X
Check that data and metadata are complete				X

6.4. Roles and Responsibilities

The importance of planning for quality in data and information before a project begins is critical. Quality assurance methods should be in place at the inception of a project and continue through all project stages to final archiving of the data set. All network employees from the network coordinator to the data entry technicians, not only the data manager, should take pride in data quality. People are the most important factor in the data quality process, and everyone plays a part in achieving high quality data products. All employees assigned to a project are responsible for the quality of the results generated from his or her task(s).

As the director of the SFAN I&M program, the network coordinator must recognize the importance of data quality and support all efforts to achieve the highest possible quality in the data produced by the program. The park data managers are responsible for:

1) developing protocols and SOPs to ensure data quality,
2) making project managers, technicians, etc., aware of the established procedures and enforcing adherence to them,

3) evaluating the quality of all data and information against NPS standards before dissemination outside the network or archiving, and
4) performing periodic data audits and quality control checks to monitor and improve the data quality program.

Project managers must:

1) be aware of quality protocols and convey their importance to technicians and field crews,
2) ensure compliance with the protocols,
3) validate data after the verification process is complete, and
4) review all final reports and information products.

Technicians must follow established protocols for data collection, data entry, and verification established in the inventory and monitoring protocol data management SOPs.

6.5. Goals and Objectives

The overarching goal in establishing goals, objectives, and criteria for data quality is to ensure that a project produces data of the right type, quality, and quantity to meet the project objectives and the user's needs. Quality criteria should be set at a level proportionate to the project-specific objectives, and these criteria should indicate the level of quality acceptable for the final data product. The EPA (2003) defines data quality objectives as qualitative and quantitative statements that:

- clarify the intended use of the data,
- define the type of data needed to support the decision,
- identify the conditions under which the data are to be collected, and
- specify tolerable limits on the probability of making a decision error due to uncertainty in the data.

The most effective mechanism for ensuring that a project produces data of the right type, quality, and quantity is to provide procedures and guidelines to assist the researcher in accurate data collection, entry, and validation. Therefore, a comprehensive set of SOPs and data-collecting protocols for quality control, namely clear field methodologies, a well-trained field staff, well-organized field forms, and data entry applications with simple built-in validation will be written.

Although specific QA/QC procedures will depend upon the individual vital signs being monitored and must be specified in the protocols for each monitoring vital sign, some general concepts apply to all network projects. The general QA/QC procedures presented in this plan were primarily adapted from the Draft Data Management Protocol (Tessler & Gregson 1997) and the ideas contained in Michener and Brunt (2000). These general guidelines will ensure that all data collected are checked for integrity before being integrated into the monitoring program databases. Refer to SOPs and monitoring protocols for specific QA/QC procedures.

6.6. Data Collection

Careful, accurate recording of field observations in the data collection phase of a project will help reduce the incidence of invalid data in the resulting data set. Unlike a typographical error that occurs when a recorded observation is incorrectly transferred from a paper field form to a digital database, an incorrect entry in the field cannot be easily corrected. Therefore, attention to detail during data collection is crucial to overall data quality and will reduce the overall frequency and criticality of errors at subsequent stages in the data lifecycle.

Paper field notebooks or data forms have been the primary methods for ecological data collection for many years. Although paper may have advantages in terms of longevity and ease of use, it does not work well under some environmental conditions, and processing options are limited until the data are transferred to digital format. As an alternative to paper, several options for electronic data collection in the field are now available, including handheld computers, automated data collection instruments, and tape recorders. Regardless of the collection method, data should ideally be transferred from one form to another only once because each transfer has the potential to introduce additional errors into the data set. One transfer should result in fewer errors, provided that appropriate QA/QC measures are incorporated into the process.

Before the data collection phase of a project begins, the park data manager is responsible for providing the protocols/SOPs for data collection and storage to the project manager. All field sheets and field data recording procedures must be reviewed and approved by the data manager and documented in the protocol SOPs. The project manager, in turn, will ensure that field crews understand the procedures and closely follow them in the field. If training is necessary, the data manager will work with the project manager to provide that training. Field technicians are responsible for proofing raw data forms in the field, ensuring their readability and legibility, and verifying and explaining any unusual entries. They are expected to understand the data collection forms, know how to take measurements, and follow the protocols.

6.6.1. Suggested Methods

The following QA/QC measures are recommended to minimize error in data collection in the field:

- Use a formatted, project specific datasheet as opposed to a field notebook. To minimize error all data must be gathered in a format that can be directly entered into a computer. Standardized data sheets effectively identify what pieces of data are to be gathered and can display data for efficient computer entry. When electronic data collection devices are not used, data will be recorded on paper data forms. Some circumstances may warrant the use of paper and writing implements that can withstand moisture, dust, and other extreme environmental conditions. Standardized data sheets that identify the pieces of information to be recorded and display the data for efficient computer entry (i.e., reflect the design of the data entry interface) will help ensure that all relevant information is recorded and subsequent data entry errors are minimized. Data sheets should contain as much basic preprinted project information as possible and sufficient space for recording relevant header data such as date, collectors, weather conditions, etc. They should clearly

specify all required information, using examples where needed to ensure that the proper data are recorded. All information added to the data sheet must be printed and clearly legible. If alterations to the information are necessary, the original information should be crossed out with a single line and the new information written next to the original entry. Information should never be erased and old information should not be overwritten. Upon return from the field, copies of all original data sheets should be made and checked for legibility and completeness (i.e., no data cut off at the edges). The copies of the data sheets will be stored as specified in the protocol SOP, and the original data sheets will be used for data entry.

- Use a hand held computer for data collection when possible. The use of handheld computers minimizes the need for manual data entry from field forms and associated transcription and data entry errors. Pendragon 4.0 software is used to develop handheld data entry forms for field use that are customized to project requirements. HotSync actions between the handheld and the desktop automatically import digital field data into project databases. A customized handheld data entry application has the advantage of incorporating on-the-spot QA/QC checks, so this data collection method probably provides the highest quality data when combined with point-of-entry data quality checks. These portable units, however, are subject to environmental constraints such as heat, dust, and moisture. Protective waterproof cases should be used in adverse weather conditions. When handheld computers are used for data entry in the field, the data must be downloaded daily to avoid potential loss of information. Thus, if a handheld unit fails during data collection, only the current day's data are lost. Batteries should be checked prior to a data collection trip, and they should be charged at the end of every field day. The use of a memory card that will store the data in case of damage to the unit or battery failure is suggested. Finally, in case the unit becomes inoperable in the field, printed data sheets should always accompany field teams on data collection trips.

- Use automated data loggers where appropriate. Instruments with their own data acquisition systems may be used to collect some types of data, such as water and air quality data. These devices can be calibrated and programmed to automatically record data and store them for later download directly to a computer, thereby eliminating the possibility for manual data entry errors. Data loggers are an efficient method for recording continuous sensor data, but routine inspections are necessary. Again, environmental constraints, as well as power (e.g., sufficient battery charge) and maintenance requirements, must be considered in the use of these instruments. Regular downloads are required since physical memory is usually limited, but the elimination of manual data entry eliminates that source of errors.

- Use a handheld tape recorder. Another alternative to paper field data forms is a handheld micro cassette tape recorder. Recorded observations are subsequently transcribed to paper or directly entered into computer files. As with other technological solutions, there are drawbacks including battery and tape maintenance, low environmental tolerance, and risk of failure. However, if a single data collector is in the field, tape recorders can provide an easily operated, high quality, efficient method of collecting data.

- Consider maintenance, calibration, and minimum timing of field equipment. Accurate field measurements are only possible if field equipment is regularly maintained and calibrated. Where appropriate, consult reference manuals for recommended maintenance and calibration procedures. Once in the field, allow time for field equipment to adjust to its environment and to take accurate measurements. Such is the case for water quality probes and GPS units. Researchers should consider maintaining records of equipment calibration and failures that accompany their field data whenever possible.

- Be organized and keep a log. Organization is the key to good data collection methods. Keeping a log of any decisions made and events that occurred will help clarify information and contribute to an accurate report.

- Ensure that field crews receive proper training. Although protocols and SOPs are in place, they cannot guarantee that high quality data will be collected. Prior to routine data collection for a project, training sessions to ensure that field personnel have a clear understanding of data collection procedures described in the SOPs should be conducted. A training program may also include a process to certify that field staff understand and can perform the specified data collection procedures. The development of a training manual may be considered for long-term monitoring data collection efforts and those that will involve a large number of field staff. Palmer and Landis (2002) provide an outline for a training manual and suggestions for planning training sessions.

- Perform quantitative assessments of data quality. A quantitative assessment of data quality during data collection activities may be performed to determine if measurement protocols are being followed and quality objectives are being achieved. The repeating of a measurement or a re-measurement is the primary tool for performing quantitative assessments. Project managers should periodically review the work of field technicians to ensure that their work does not drift from standards during the course of the field season. Quantitative assessments may be considered if staff and funding are available, and Palmer and Landis (2002) describe several approaches that may be employed.

6.7. Data Entry

Data entry is the initial set of operations where data written on paper field forms are transcribed, or typed, into a computerized form linked to database tables. Spreadsheets should **not** be used for data entry (data can be exported to a spreadsheet for manipulations post entry). Where data were gathered and/or stored digitally in the field (e.g., on a datalogger or Global Positioning Systems (GPS) device), data entry is the transfer of data (downloading) to a file in an office computer where they can be further manipulated.

Superficially, getting data from field projects into the computer seems to be a fairly simple task – the process of typing it in. Nevertheless, data entry is not a trivial concern because the value of the data depends upon their accuracy. Without proper preparation and some established guidelines, the quality and integrity of the data will be debatable. Data entry is best performed by a person who is familiar with the data and ideally takes place as soon as data collection is

complete. The single goal of data entry is the *transcription* of the data from paper records into the computer with 100% accuracy. However, because transcription errors are virtually unavoidable during data entry, they will have to be corrected during the data verification process. Observation of certain data entry guidelines, however, will minimize verification work.

The data manager, in conjunction with the project manager, should provide training in the use of the database to all data entry technicians and any other users. The project manager will ensure that data entry technicians understand how to enter data and that they follow the protocols. Data entry technicians are responsible for becoming familiar with the field data forms and differences in handwriting.

The most robust QA/QC measures for data entry will be built into the database design, as described in Appendix C. Several QA/QC procedures must additionally be followed by the database user:

- Have a familiarity with the database software, database structure, and any standard codes for data entry that have been developed. Know how to open the data entry form, create a new record, and exit the database properly. Also, learn how to commit both a "field" entry and a complete record entry and correct mistakes made while typing.

- Enter or download data in a timely manner. All data should be entered or downloaded into the project database as soon as possible, preferably no less than once a week. Do not delay data entry until all the project data have been collected if at all possible.

- Enter the data, one logical "set" at a time. Record in your notebook errors you know you've made or any questions that arise about the data content; these will be useful during data verification. Initial and date each paper form as it is completed to avoid confusion about what has been entered and what has not with a different color then the data. Interrupt your data entry only at logical stopping points. If you reach stopping points, make a working backup copy of the data for safety's sake if your software does not do so automatically.

6.8. Verification and Validation Procedures

Data quality is appraised by applying verification and validation procedures as part of the quality control process. These procedures are more successful when preceded by effective quality assurance practices. Performing both verification and validation of data must be stressed because it is important to remember that verified data are not always valid data. Data verification checks that the digitized data match the source data, whereas data validation checks that the data make sense. It is essential that all data are validated as truthful and do not misrepresent the circumstances and limitations of their collection. Failure to follow SOPs for data entry, validation, and verification will render a data set suspect. It is important to remember that only the data entry and verification stages can be done by someone who is not familiar with the kinds of errors sought during validation; validation requires in-depth knowledge about the data. As a general rule, data in each project database will be reviewed and corrected using

approved verification and validation methods such that data accuracy is 95% or greater. Database fields should be included to record concerns about data integrity when applicable.

6.8.1. Data Verification

Manual effort is generally required to get data into electronic format. Any typographical errors made will accumulate in the permanent database unless the data are verified and the errors detected. By implementing data verification procedures, these errors can be reduced, if not eliminated. Data verification immediately follows data entry and involves checking the accuracy of the computerized records against the original source, usually hard copy field records, and identifying and correcting any errors. When the computerized data are verified as accurately reflecting the original field data, the paper forms can be archived and most data manipulation can be done on the computer.

6.8.2. Suggested Methods for Data Verification

Each of the following methods has a direct correlation between effectiveness and effort. The methods that eliminate the most errors can be very time consuming while the simplest and cheapest methods will not be as efficient at detecting errors.

- Visual review at data entry. The data entry technician verifies each record after it is input. The values recorded in the database are compared with the original values from the hard copy and any errors are corrected immediately. This method is the least complicated since no additional personnel or software is required. The reliability of this method depends wholly on the person keying data and is generally the least reliable of the data verification methods.

- Visual review after data entry. Visual review after data entry requires two people. One person sits with the field datasheets in hand while the other sits at the computer with the database open. As data sheets are read allowed, the database record is followed and analyzed for data entry errors. Errors are corrected as they are discovered. Verified data sheets should be dated and initialed. A "verified" check box accompanying each record in the database should be checked when the record is verified, no matter which method is used.

In addition to the above methods, simple summary statistics using the entered data can be calculated with statistical software. This is important because even when care is taken up to this point, a duplicate or omitted entry may have been overlooked. For example, the number of known constant elements, such as the number of sampling sites, plots per site, or dates per sample, can be viewed. The same question can be posed in different ways; differences in the answer provide clues to errors. The more checks devised to test the completeness of the data, the greater the confidence that the data are completely verified.

To minimize transcription errors, our goal is to verify 100% of records to their original source. In addition, 10% of records are reviewed a second time by the project manager, and the results of

that comparison are reported with the data as an additional metadata component. If errors are found in the project manager's review, then the entire data set is verified again.

6.8.3. Data Validation

Although data may be correctly transcribed from the original field forms (data entry and verification), the data may be neither accurate nor logical. For example, entries of stream temperature of 95°C in data files raise doubt about their accuracy; and such entries almost certainly are incorrect, whether or not they were properly transcribed from field forms. This process of reviewing computerized data for range and logic errors is *validation*. It can be done during data verification *only* if the operator has comprehensive knowledge about the data. More often, validation is a separate operation carried out *after* verification by a project specialist to identify generic and specific errors in particular data types. Corrections or deletions of logical or range errors in a data set require notations in the original paper field records about how and why the data were changed. Modifications of the field data should be clear and concise while preserving the original data entries or notes. Validation efforts should also include a check for the completeness of a data set because field sheets or other sources of data could easily be overlooked.

General step-by-step instructions are not possible for data validation because each data set has unique measurement ranges, sampling precision, and accuracy. Nonetheless, validation is a critically important step in the certification of the data. Invalid data commonly consist of slightly misspelled species names or site codes, the wrong date, or out-of-range errors in parameters with well defined limits (e.g., elevation). But more interesting and often puzzling errors are detected as unreasonable metrics (e.g., stream temperature of 95°C) or impossible associations (e.g., a tree 2 feet in diameter and only 3 feet high). These types of erroneous data are called *logic errors* because using them produces illogical (and incorrect) results. The discovery of logic errors has direct, positive consequences for data quality and provides important feedback to the methods and data forms used in the field. Validation, therefore, is not a step to be ignored until after statistical analyses reveal problems with the data. Project specific data validation procedures will be described in the project specific protocols.

6.8.4. Suggested Methods for Data Validation

Wherever possible, the data entry application should be programmed to do the initial validation using database design standards outlined in Appendix C. Certain components of data validation are built into data entry forms. The simplest validation during data entry is range checking, such as ensuring that a user attempting to enter a pH of 20.0 gets a warning and the opportunity to enter a correct value between 1.0 and 14.0 (or better yet, within a narrow range appropriate to the study area). Not all fields, however, have appropriate ranges that are known in advance, so knowledge of what are reasonable data and a separate, interactive validation stage are important.

Additional data validation may be performed via exploratory data analysis to look for outliers. Database, graphic and statistical tools can be used for ad-hoc queries and displays of the data. Such exploratory techniques will help identify obvious outliers.

Data quality assurance procedures should not aim to eliminate outliers. Extreme values naturally occur in many ecological phenomena; eliminating these values simply because they are extreme is equivalent to pretending the phenomenon is 'well-behaved' when it is not. Eliminating data contamination is a better way to explain this quality assurance goal. If contamination is not detected during data collection, it is usually detected later if an outlying data value results. When an outlier is detected, attempts should be made to determine if some contamination is responsible.

With a catalog of errors and some exploratory data results in hand, it makes sense to reevaluate the field data forms as the source of the logic errors. Often minor changes (or small font annotations) to a field form will remove any ambiguity regarding what to enter on spaces in the form. In fact, any time the same type of validation errors occur repeatedly in different datasets, it is usually the field form that is at fault, not the field crew. Of course, it could also mean that the protocol or the field training is faulty.

6.9. Version Control

The SFAN manages files from a multitude of sources, comprised of many formats with many iterations of a particular product. Some of the files are complete, some are works-in-progress, and for others the status cannot be determined. In addition to files collected by the network, the network has also generated many files, some of which fall into the complete, works-in-progress, and undetermined status categories. Determining the status of a single file can be difficult, but when confronted with a series of similarly named files, the task of determining which is the most current can become impossible.

Version control is the process of documenting the temporal integrity of files as they are being changed or updated. Change includes any alteration in the structure or content of the files, and such changes should not be made without the ability to undo mistakes caused by incorrect manipulation of the data. Data progresses through various lifecycle stages, and whenever a set of changes is complete, the user should save the file with a unique name. Version control is simple insurance for maintaining data integrity, and using good version control should be routine for all data handlers.

Prior to any major changes to a file, a copy should be stored with the appropriate version number that allows the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used for data entry or analysis.

6.9.1. Version Control Options

Dates. Using a date provides logical version control. The date is usually formatted as YYYYMMDD or YYMMDD, where DD is optional (depending on the frequency of changes).

One drawback to this method is that dates may be hard to read, thus causing confusion for users who may open the wrong version of a file.

Sequential numbers. Versioning of archived data sets is handled by adding a number to the file name, e.g., 001 or V1.0 for the first version. Each additional version is assigned a sequentially higher number. Documenting the date that a file becomes a new version is strongly recommended if this method is used. For example, backup copies of the same database with different raw data are placed in a backup subdirectory with the YYYYMMDD date tagged on the end. Frequent users of the data must be aware of the version control method so they can identify the most recent version.

6.9.2. Version Control Documentation

Databases that are converted from one version of MS-Access to an upgraded version will require additional QC; in particular, if the databases are being actively used for data entry or analysis. Forms, queries, reports, and data entry all will be thoroughly tested. Data dictionaries that fully document field names and relationships should be defined. In addition table keys and constraints must be created each time there is a significant change to the database. Data dictionaries and data management SOP's from previous versions should be archived along with the original databases.

6.10. Data Quality Review and Communication

QA/QC review is required prior to communicating/disseminating data and information. Only data and information that adhere to NPS quality standards will be released.

Director's Order #11B states that all information (e.g., brochures, research and statistical reports, policy and regulatory information, and general reference information) distributed by the NPS (including information obtained from sources outside of the NPS) must be accurate, reliable and timely in nature. Therefore, the SFAN must evaluate and identify the types of information it will disseminate that will be subject to the guidelines. Information disseminated to the public must be approved by the appropriate reviewing officials and programs. Documentation of the QA/QC standards used in producing the information and that substantiate the quality of the information must be formally documented. Furthermore, mechanisms must be in place for receiving and addressing comments/complaints pertaining to the quality of data.

Data are distributed to the public through the SFAN I&M web page, national web sites such as the Biodiversity Data Store and the Natural Resource/GIS Data Store, and through public access databases such as NPSpecies and NatureBib. Any information distributed through any of these mechanisms must undergo internal QA/QC procedures and be approved for release.

6.10.1. Data Quality Review Methods

The SFAN Inventory and Monitoring Program will establish guidelines and protocols to ensure compliance with DO #11B. These protocols will document both internal and external review

procedures for data and information disseminated outside the network, as well as a process for processing complaints about data quality.

6.10.2. Value of Feedback from QA/QC Procedures

Quality assurance procedures may need revision to improve the quality level if random checks reveal an unacceptable level of data quality. Quality checks should not be performed with the sole objective of eliminating errors; the results may also prove useful in improving the overall process. For example, if the month and day are repeatedly reversed in a date field, the data entry technicians may require retraining about the month/day entry order. If retraining is unsuccessful in reducing the error's occurrence, the computer program may need to be rewritten so that month and day are entered separately, field length limits are enforced, or a pick list is created. In this manner, the validation process will serve as a means of improving quality as well as controlling the lack of quality.

Sometimes, modification of field data forms to avoid common mistakes is necessary. With knowledge of validation errors and exploratory data results in hand, the field data forms as the source of the logic errors can be reevaluated. Often minor changes, small annotations, or adding check boxes to a field form remove ambiguity about what to enter on the form. In fact, any time the same type of validation error occurs repeatedly in different data sets, the field form – not the field crew – is usually at fault. Repeated errors found during validation can also mean that protocols or field training are at fault, which can then be recognized and corrected.

6.10.3. Monitoring Conformance to Plans and Standards

Data managers may elect to use periodic data audits and quality control checks as mechanisms to actively participate in the oversight and improvement of a data quality program. Data managers must verify that staff is operating in conformance with the data quality procedures specified in this plan and the protocol specific data management plans. The data manager should track and facilitate the correction of any deficiencies. These quality checks promote a cyclic process of continuous feedback and improvement of both the data and quality planning process.

Periodic checks by the data manager to see if network staff are adhering to the data quality procedures established in the Data Management Plan and protocols SOPS may include verification of the following:

- Data collection and reporting requirements are being met.
- Data collection and reporting procedures are being followed.
- Verification and validation procedures are being followed.
- Data file structures and maintenance is clear, accurate and according to plan.
- Revision control of program documents and field sheets is adequate.
- Calibration and maintenance procedures are being followed.
- Seasonal and temporary staff have been trained in data management practice.
- Metadata collection and construction for the program is complete.
- Data is being archived and catalogued appropriately for long term storage.

The results of quality assessments should be documented and reported to the research staff and the network coordinator. The project manager and coordinator are responsible for ensuring that non-conformities in data management practices are corrected.

6.10.4. Communicating Data Quality

Data documentation and metadata will be used to notify end users, project managers, and network management of data quality. A descriptive document for each data set/database will provide information on the specific QA/QC procedures applied and the results of the review. Descriptive documents or formal FGDC-compliant metadata will document quality for spatial and non-spatial data files posted on the Internet.

7. Data Documentation

7.1. Purpose of Metadata

Data documentation is a critical step towards ensuring that data are useable for its intended purposes well into the future. This involves the creation of metadata. Metadata can be defined as data about the content, quality, condition and other characteristics of data. Additionally, metadata provide the means to catalog datasets, within intranet and Internet systems, thus making these datasets available to a broad range of potential data users.

Data sets seem to sometimes take on lives of their own. Some seem to have the ability to reproduce and evolve on multiple hard drives, servers and other storage media. Others are masters at remaining hidden in digital formats or in forgotten file drawers. In addition, once data are discovered, a potential data user is often left with little to no information regarding the quality, completeness, or manipulations performed on a particular "copy" of a dataset. Such ambiguity results in lost productivity as the user must invest time tracking information down, or, worst case scenario, renders the dataset useless because answers to these and other critical questions cannot be found. As such, data documentation must include an upfront investment in planning and organization.

While the importance for metadata is universally accepted within the data management community, the approaches for collection and levels of detail are varied. The following are some considerations which an I&M network or prototype manager should consider in the development of data documentation strategies.

- Executive Order 12906, signed by President William Jefferson Clinton in 1994, mandates federal agencies to "...document all new geospatial data it collects or produces, either directly or indirectly..." using the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM). In addition, EO 12906 directs agencies to plan for legacy data documentation and provide metadata and data to the public.
- The FGDC Biological Data Profile contains all the elements of the CSDGM and includes additional elements for describing biological data sets. Metadata created in compliance with the Biological Data Profile can be added to the National Biological Information Infrastructure (NBII) Clearinghouse. Although not a requirement, completion of the Biological Data Profile for appropriate data sets is recommended.
- All GIS data layers must be documented with applicable FGDC and NPS metadata standards. The NPS GIS Committee requires all GIS data layers be described with FGDC standards and the NPS Metadata Profile.
- While there are numerous tools available for collecting metadata, the NPS Integrated Metadata System Plan is limited to three recommended desktop applications: Dataset Catalog, ArcCatalog, and Spatial Metadata Management System (SMMS).

7.2. NPS Integrated Metadata System Plan and Tools

As noted above, the NPS Integrated Metadata System Plan is limited to three recommended desktop applications for collecting metadata. These include Dataset Catalog, developed by the I&M Program, and two commercial off the shelf (COTS) metadata tools, ArcCatalog and SMMS. A brief description of each of these tools, including their potential utility in metadata creation follows below. A fourth tool, the Metadata Parser (mp) and its utility in metadata creation is also briefly discussed.

Dataset Catalog: Dataset Catalog is a tool for cataloging abbreviated metadata on geospatial and biological data sets pertaining to park(s) and/or a network. It provides parks and/or networks a means whereby they can inventory, organize, and maintain information about data set holdings locally. While Dataset Catalog is not intended to be an exhaustive metadata listing, it does assist parks and networks in beginning to meet the mandates of EO 12906. With the current version of Dataset Catalog, (version 2), records can be exported as an FGDC text file, which can than be imported into other metadata tools. The I&M Program recommends all relevant datasets at I&M parks and networks be cataloged in at least simple Dataset Catalog format.

Spatial Metadata Management System: SMMS is a tool with the capability to create, edit, view, and publish metadata that is compliant with FGDC requirements. SMMS uses an MS Access database structure combined with an advanced FGDC-compliant metadata editor. The software allows selection of views depending on whether the user wants the full standard, biological, or the minimal compliant view of FGDC Sections 1 and 7. There is online Help to describe the purpose, usage or mandatory status of metadata elements. The context-sensitive help file provides the FGDC definition for each field on the screen. In addition to Help files, there are sample metadata records for most sections that provide "real world" examples. The NPS Integrated Metadata System Plan recommends SMMS for FGDC Biological Profile and other geospatial metadata creation.

ArcCatalog: ArcCatalog is a management tool for GIS files contained within the ArcGIS Desktop suite of applications. With ArcCatalog, users can browse, manage, create, and organize tabular and GIS data. In addition, ArcCatalog comes with support for several popular metadata standards, including the FGDC standard, that allow one to create, edit, and view information about the data. There are metadata editors to enter documentation, a storage schema, and property sheets to view the metadata. With ArcCatalog users can view GIS data holdings, preview geographic information, view and edit metadata, work with tables, and define the schema structure for GIS data layers. Metadata within ArcCatalog is stored exclusively as Extensible Markup Language (XML) files. The NPS Integrated Metadata System Plan recommends ArcCatalog for gathering GIS-integrated geospatial metadata.

An optional but highly recommended extension for ArcCatalog is the NPS Metadata ArcCatalog Extension developed by NPS Midwest Region GIS Technical Support Center. The extension fixes several ArcGIS 8 metadata errors and provides added functionality for NPS users. Development is also underway to provide Biological Profile editing capability and NPS Profile support.

Metadata Parser: The MetaParser (mp) program is used to validate metadata records by checking the syntax against the CSDGM and to generate compliant output files for posting to clearinghouses. It generates a textual report indicating errors in the metadata, primarily in the structure, but also in the values of some of the scalar elements where values are restricted by the standard.

Until recently many NPS data stewards collected, parsed and stored metadata (and GIS data sets) in the NPS GIS Clearinghouse managed by North Carolina State University (NCSU). However, efforts are currently underway to unify and streamline metadata development. This new approach utilizes existing desktop metadata creation applications, as well as an online integrated metadata database (NR-GIS Metadata) and a web based data server (NR-GIS Data Server). NR-GIS Metadata and NR-GIS Data Server will comprise a web based system to integrate data dissemination and metadata maintenance. It will be possible to update Dataset Catalog records in the NR-GIS Metadata database or in the source desktop application (i.e., ArcCatalog, Dataset Catalog, SMMS). Non-sensitive NR-GIS Metadata records are automatically posted to NPSFocus.

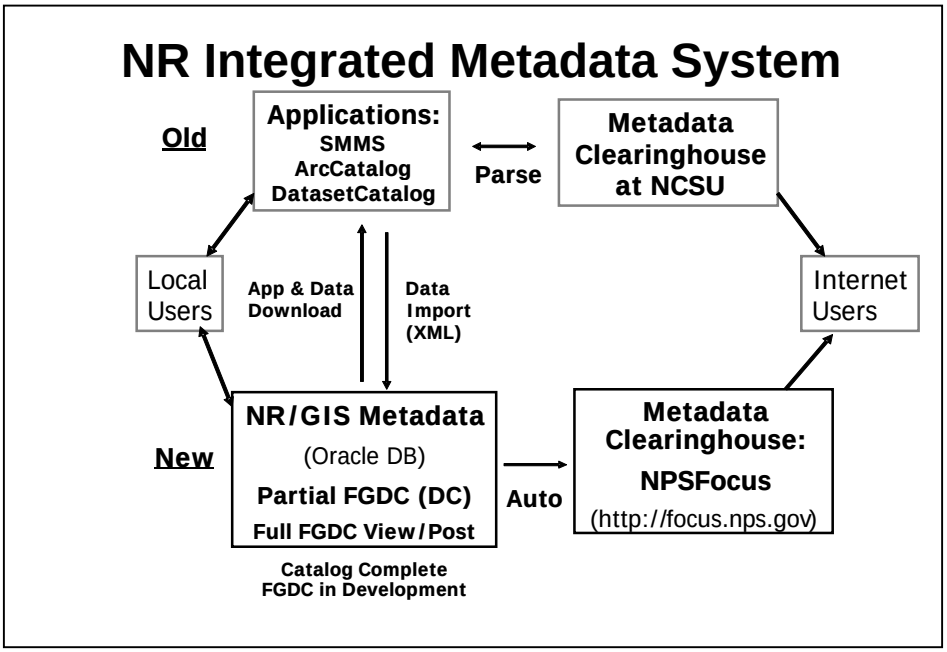


Figure 7.1. Natural resources integrated metadata system.

7.3. Metadata Process/Workflow

SFAN data managers and GIS specialists currently oversee metadata creation for I&M project data sets. In the future, the data management team intends to train project managers and technicians to create their own metadata, or at least make significant steps towards FGDC compliance with their metadata. In a formal protocol documenting the metadata creation process, the SFAN will:

- review FGDC required fields following the CSDGM, using guidance from the FGDC Content Standards for Digital Geospatial Metadata Workbook Version 2.0 (2000)
- provide step by step instruction in the creation of metadata using ArcCatalog, SMMS, and Dataset Catalog
- provide information on how to create metadata templates in these programs
- provide guidance on how to post spatial data, tabular data, and associated metadata to the Biodiversity Data Store and NR-GIS Metadata and Data Store

The SFAN metadata protocol will be posted to the SFAN website and provided to contractors, who will be required to complete full FGDC compliant metadata on any spatial and tabular datasets submitted to the SFAN.

Until the SFAN metadata protocol is developed, the following will serve as the basic steps that the SFAN I&M data management team uses to create metadata for I&M data products:

Step 1: Spatial and Non-Spatial Data

Identify Relevant Data Sets and Compile Pertinent Metadata

Data utilized by the Network can be grouped, at least initially, in three broad categories based on origin. These categories include legacy datasets (primarily data collected prior to the inception of the Network), non-programmatic datasets (ongoing data collection efforts conducted by entities outside the direct purview of the Network), and programmatic datasets (data collected by, and/or under the direct purview of, the Network).

Legacy Data
In many cases, legacy data are initially identified as part of data mining efforts. Unfortunately, many of the legacy datasets will be missing pertinent information, and the originator may no longer be in contact. Thus, an "adequate" level of documentation may not be possible. However, the data and all supporting documentation related to it should be assembled and reviewed. Many legacy data sets will need to be converted to a standard database format for incorporation and future analyses. Data entry, validation and verification procedures will follow those contained within this Data Management Plan. A processing and revision log in text format will be maintained with the dataset to help develop pertinent metadata.

Non-programmatic Data
Networks and/or prototypes are not the only entities gathering relevant inventory and/or monitoring data pertinent to park management. The Network and Prototype will make every effort to capture and assimilate all relevant data. The outside entity will be contacted and a request will be made for available metadata and/or a metadata interview will be conducted. Similar to legacy data, data may need to be converted to a standard database format for analysis. Data entry, validation and verification procedures will follow those contained within this Data Management Plan. A processing and revision log will be maintained with the dataset for capture of additional metadata.

Programmatic Data

For new projects, metadata consideration will begin up front, with an interview with the principal investigator to explain what will be needed to properly document the data. In most instances, this will include completion of a basic metadata survey for inclusion in the data manager's project file, as well as submission of supporting documentation (proposal, SOPs, etc.). In addition, a database structure will be developed by, or in close consultation with, the data manager, to ensure compliance with the principles and procedures contained within this DMP. Updates/revisions to the metadata will be conducted in tandem with data submissions.

Step 2: Non-Spatial Data

Create Dataset Catalog Record

The Network will develop a Dataset Catalog record for relevant non-spatial data, including inventory datasets and monitoring project databases. The "standard" view of the Dataset Catalog data entry form will be used in order to document data tables and field descriptions within project databases.

Prioritization of datasets for further documentation will be based upon current or anticipated future use. In other words, datasets which will be used repeatedly in analysis or with high probability for data sharing will be addressed first.

Step 2: Spatial Data

Select Metadata Tool and Begin Record

Within the SFAN, GIS and data management personnel currently utilize both SMMS and ArcCatalog to generate FGDC compliant metadata for spatial data. Both software programs are acceptable means to this end and will continue to be used. Metadata templates will be established to reduce the effort required to complete full FGDC compliant metadata for spatial data.

Step 3: Spatial Data

Import into Dataset Catalog and Complete Record

A useful tool in Dataset Catalog is the ability to import metadata text files for inclusion into the database. Spatial metadata records will be exported from SMMS or ArcCatalog as text files and then imported into Dataset Catalog. Additional fields specific to Dataset Catalog, such as NatureBib reference numbers, will be added after metadata text files have been imported.

This approach provides metadata for all Network data holdings, both spatial and non-spatial, in a searchable, centralized location. These records can be imported into the online NR-GIS Metadata System or continue through additional processing steps based on data type, source, and importance.

Step 4: Spatial and Non-Spatial Data

Make Information Available

At a minimum, metadata and associated data should be submitted to NR-GIS Metadata and Data Store. This can be accomplished utilizing the recommended desktop applications. Additionally, information on data holdings should be conveyed in a meaningful manner for park resource managers, researchers, and others with a potential interest/stake in park management and/or research endeavors. Similar to metadata creation the mechanisms and formats for accomplishing this are varied. In addition to FGDC Text files, Dataset Catalog can output a list of all records, single record reports, and/or a data dictionary report. Dependant upon the target audience, these standardized outputs can be useful in conveying information on program data holdings and summaries of database structures. Obviously, customized queries and reports can also be generated. Other standardized outputs include ArcCatalog stylesheets. The NPS Metadata ArcCatalog Extension contains custom stylesheets, which can be invoked from the metadata toolbar. These can be utilized to depict pertinent details in a more coherent format than standard metadata outputs.

7.3. Additional Database Documentation

In additional to developing compliant metadata for project databases, data managers will detail project database descriptions in project SOP's. Database documentation in this format can greatly expand on metadata documentation by displaying table relationships, illustrating key data entry forms, and outlining automatic reporting procedures. While metadata accurately describes data components such as data quality and field definitions, metadata generally does not provide an opportunity to describe how a database actually functions.

Database documentation contained within project SOP's will help to train data entry technicians, assist with the peer-review of data management practices for specific projects, and facilitate the transfer of data management responsibilities in the case of a change in personnel.

8. DATA ANALYSIS AND REPORTING

As described in the previous chapter, sound data management practices are the key to having a credible monitoring program that provides useful data to managers. Monitoring data must be analyzed, interpreted, and provided to managers and other decision-makers and interested parties in a useable form at regular intervals if a monitoring program is to be successful. Different types of reports are needed to provide information to multiple audiences.

This chapter presents an overview of the types of analyses and reporting envisioned for the SFAN I&M program. The protocols included in this discussion are those on the "implementation list"; that is, those protocols whose implementation we will fund in the near-to mid-term. The closer a protocol is to completion, the more we know about that program, what data will be collected, what analyses will be performed, what products will be generated, to what end and for what audiences. Detailed information on data analysis will be contained in the full monitoring protocols.

Data analysis and reporting must be incorporated into the annual project timeline and given equal importance to actual data collection. Project timelines, including reporting cycles, for funded monitoring programs are presented in Appendix F.

8.1. Data Analysis

Identifying project objectives prior to commencing monitoring activities is the first step in data analysis and product development. Monitoring databases will generally come directly from Project Leaders and will be as complete as possible prior to analysis. Weather data, complete site information, taxonomic information, complete names of observers, final data verification, and even database design changes such as keys, indexes, and constraints should be complete prior to analysis. The development of data products will be guided by the Vital Sign project objectives, and when available, by the protocols and data management and data analysis SOPs.

Summary analysis for annual reports of monitoring studies should include relevant descriptive statistics (mean, standard deviation, sample size) for all primary variables included in the project. Indices for calculating species richness, diversity and evenness for plant and animal community data can also be useful.

Two factors permeate all data analysis and place constraints on subsequent biological interpretation of field data and management recommendations derived from such interpretation. First, field data are samples and statistics. Derived from them are estimates of 'real' but unknown dimensions or relationships. Second, when evaluating comparable statistics from different populations using a statistical test, the power of the test must be understood. In essence, it is necessary to know the ability of the test to detect a difference, given that the difference actually exists. Thus, the guiding approach towards data analysis and interpretation of results must utilize appropriate methods. These will include tests of statistical significance along with associated power analyses.

Further details of data analyses, especially in regards to trend detections, are program specific and are described in detail in respective program SOPs.

8.2. Data extraction

8.2.1. Identifying required tables and fields

Project scopes-of-work, and whenever possible, annual reports from previous years will be used to identify the tables and fields necessary for data summary. Assuming a database already exists, one can begin by identifying the core or critical tables with the desired data addressing the project goals. The total database design, including relationship diagrams, priority fields, and the tables that contain them, will additionally be reviewed to work towards building queries for data summary and analysis. These are also the preliminary steps that will be used for report automation.

8.2.2. Combining tables to extract the necessary fields

Individual fields will be identified which provide the source data for the desired analyses. Very often, the fields are to come from two or more tables, so it will become necessary to join tables using queries. Relationship diagrams will be used to determine the most effective way to write queries. Usually analysis queries do not exist. The Data Manager in coordination with the Project Manager will build these queries. The Project Manager will start by identifying one or two variables to be used in the analysis, and then add independent variables that will be used for grouping purposes such as site, habitat, or time interval. These will form the basis for the analysis queries. The analysis will be developed through a series of queries until either the data are exported to outside applications or the desired result is calculated.

8.3. Automated Reporting and Data Summaries

The automation of data summaries and annual reports facilitates the SFAN's ability to manage multiple projects. The Network uses Microsoft Access to automate its data summaries and reports within project databases.

8.3.1. Annual Report as Template

The development of automated reporting would be greatly facilitated by existing annual reports when these are available. If not, a detailed annual report template will be worked out between the Project Manager and the Data Manager. The Data Manager will take the extracted tables and create an output that is similar, if not identical, to the existing annual report template. This task requires that the Project Manager have a very clear idea of what will be needed for the annual reports. Usually, the Data Manager will need only to develop the necessary queries, macros, modules and reports. The process of automation is then to link the various database objects to a simple user-interface.

8.3.2. Project Manager Interviews

Often, the Project Manager will need to modify the existing annual reports to better meet the requirements of Vital Signs projects. If possible, the Data Manager will plan on expanding and improving the analysis and reporting products during the automation phase. Developing the automated reporting system beyond the existing annual reports will require frequent and repeated interviews by the Data Manager with the Project Manager. Once again, the project objectives will facilitate communication between the Data Manager and the Project Manager.

8.3.3. Implementing the Automated Report

Following discussions on data summary and reporting requirements, the next step is to implement the automated report. Except in the simplest of situations, the SFAN will develop each analysis independently through a series of Access queries. The resultant query will be loaded as one of several controls (textboxes) into the summary report. All queries that feed into a single report may be triggered from a form button, a macro, or a combination of both. The Network will use modules and VBA for automation only to the extent that more complex automation applications are required.

8.4. Exports to statistical packages and other software

At times, data will need to be exported out of the database to other software applications. The Network is planning to export data from Access databases for statistical analysis beyond means, standard deviations, and other descriptive statistics. The Network will be using third party statistical software for frequency distribution plots, tests for normality and analysis of variance such as SAS and Statview.

8.4.1. Field order, field type and analysis objectives

Analysis objectives will be reviewed by the Data Manager and Project Leader together prior to attempting data exports. The field order (order of variables in the resulting 'flat' file) is most easily controlled while the data are still in the relational database. Therefore, the Data Manager will assist in setting up the necessary queries to provide the field order. Field types are typically determined early on, when the database is being designed. To minimize risk of data loss as a result of data type conversion, changes in data types will be kept to a minimum once monitoring projects are underway.

8.4.2. ASCII text format

ASCII text has the advantage of being almost universally readable by third party applications. Typically, fields are delimited by commas, tabs, or spaces. Text strings are usually enclosed by single or double quotes. ASCII text requires extra steps transferring data between applications and also extra care in formatting the data. The Network will be using ASCII format only when other options are not available. The preferred alternatives are data exports directly from an

Access .mdb file or link to an .mdb file by way of a database connection such as OLE DB or ODBC data link (many Windows-based applications).

8.5. Data Reporting

The range of analysis and reporting requirements for the I&M program include:

- Annual reports
- Analysis and synthesis reports
- Program and protocol reviews
- Scientific journal articles and book chapters
- Interpretation and outreach

8.5.1. Annual Reports

The major purpose of annual reports is to:

- archive annual data and document monitoring activities for the year
- describe current condition of the resource
- document changes in monitoring protocols
- increase communication within the park and network.

Many of our monitoring programs will be active each year, and those programs will generate annual reports each year. However, some programs will be reviewed to see if costs savings may be accrued by reducing sampling effort without losing significant predictive power. If some protocols begin to collect data every other year, every 4th year, or even every 10th year, those programs will produce "annual" reports only during those years where there are significant monitoring activities to document.

8.5.2. Analysis and Synthesis Reports

The role of analysis and synthesis reporting is to:

- determine patterns/trends in condition of resources being monitored
- discover new characteristics of resources and correlations among resources being monitored
- analyze data to determine amount of change that can be detected by this type and level of sampling
- provide context, interpret data for the park within a multi-park, regional or national context
- recommend changes to management of resources (feedback for adaptive management).

These reports can provide critical insights into resource status and trends, which can then be used to inform resource management efforts and regional resource analyses. This type of analysis, more in depth than that of the annual report, requires several seasons of sampling data. Therefore, these reports are not written more frequently than every three to five years, for resources sampled annually. For resources sampled less frequently, or which have a particularly low rate of change and variability, intervals between reports may be longer. The target audience

for analysis and synthesis reports is Superintendents, park resource managers, network staff, and external scientists.

8.5.3. Program and Protocol Reviews

Purpose:

- Periodic formal reviews of operations and results
- Review protocol design and products to determine if changes needed
- Part of quality assurance – peer review process.

The primary audience for this type of review includes Superintendents, park resource managers, network staff, Servicewide Program managers, and external scientists. Protocol reviews are usually initiated by program managers and should be incorporated with or at least closely coincide in timing with the data synthesis reports described in Section 8.2.2. The Network Coordinator will initiate the Network Inventory and Monitoring Program review.

The following is an example of the sequence of events in a program review:

- Program manager/network team summarizes program and activity to date
- Conduct program assessment (e.g., power analyses of the data) and report findings
- Broad spectrum of peers invited to review the above products
- Peers invited to a workshop to discuss the program, the analyses it was subjected to, whether or not it is meeting program goals, how it might be improved, what should be changed, left behind, or tried anew, and what new partners might be enlisted to join.
- The program manager or contracted personnel compiles results from this workshop, circulates a report to participants for review, posts the final report on the SFAN web site, and finally sends all relevant final drafts to NPS regional and WASO program offices.
- The program manager develops strategies with the SFAN team on how and when to implement different recommendations.

Typical topics addressed are a general review of program efficacy, accountability, scientific rigor, contribution to adaptive park management and larger scientific endeavors, outreach, partnerships and products. These reviews are among the most in depth - not only will monitoring results be analyzed over a longer period of time, but the entire program, its structure, and function are evaluated to determine not only whether the program is achieving its objectives, but also whether the list of objectives is still relevant, realistic, and sufficient.

8.5.4. Scientific Journal Articles

This aspect of the program will be directed by the program managers, and is more at the discretion of the individual investigators than any of the previous report types. Publishing scientific journal articles is primarily conducted to communicate advances in knowledge, and is a very important, widely acknowledged means of quality assurance and quality control, via the peer review process. Putting a program's methods, analyses, and conclusions under the scrutiny of a scientific journal's peer review process is basic to science and one of the best ways to ensure scientific rigor. This peer review process is conducted by the journal editor, so no coordination at the network level is required.

Scientific journal articles produced by SFAN efforts are tracked by the SFAN monitoring program; new publications are a standard part of the Annual Administrative Report and Work Plan (see Annual Reports section), which the network updates and sends to the regional and national offices each year. Additionally, the network data managers will enter all scientific journal articles that are reported in the AARWP into the NatureBib database . Principle Investigators of recently published work in the SFAN frequently make presentations at professional workshops and conferences, and some will be invited to present their project and findings at Technical Committee and Board of Directors meetings.

8.5.5. Interpretation and Outreach

Scientific information gained from monitoring programs usually requires a concerted effort to be translated for the general public. There are efforts in several network parks which produce natural history newsletters and other media to share what has been discovered with the public. Interpretation and outreach is a perfect place for the SFAN Vital Signs Monitoring program to team up with the Pacific Coast Learning Center (PCLC). The latter program is designed to promote research in parks, as well as act as a bridge between science being conducted in parks and the public. The SFAN Network Coordinator and SFAN program leads are working with the PCLC program to form connections with college students, partners, and the interested public so that not only will information gleaned from the Vital Signs Monitoring program get out into the community, but when it arrives, it will be in a form that people can readily grasp, understand, and appreciate.

9. Data Ownership

9.1. National Park Service Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park Service.
- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.
- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, "Intangible Property" describes the following administrative requirements pertinent to data and ownership:

> *(a) The recipient* (higher institution or non-profit organization receiving federal monies for natural resource inventory and/or monitoring) *may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The Federal awarding agency(ies*) (in this case the National Park Service) *reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.*

Section 36 also states:

> *(c) The Federal Government has the right to:*
>
> > *(1) obtain, reproduce, publish or otherwise use the data first produced under an award*

(2) authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes

(d) (1) In addition, in response to a Freedom of Information Act (FOIA) request for research data relating to published research findings produced under an award that were used by the Federal Government in developing an agency action that has the force and effect of law, the Federal awarding agency shall request, and the recipient shall provide, within a reasonable time, the research data so that they can be made available to the public through the procedures established under the FOIA. If the Federal awarding agency obtains the research data solely in response to a FOIA request, the agency may charge the requester a reasonable fee equaling the full incremental cost of obtaining the research data. This fee should reflect costs incurred by the agency, the recipient, and applicable subrecipients. This fee is in addition to any fees the agency may assess under the FOIA (5 U.S.C. 552(a)(4)(A)).

(2) The following definitions apply for purposes of paragraph (d) of this section:

(i) Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This "recorded" material excludes physical objects (e.g., laboratory samples)…

(ii) Published is defined as either when:

(A) Research findings are published in a peer-reviewed scientific or technical journal; or

(B) A Federal agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

(iii) Used by the Federal Government in developing an agency action that has the force and effect of law is defined as when an agency publicly and officially cites the research findings in support of an agency action that has the force and effect of law.

9.2. Establishing Data Ownership Guidelines

The San Francisco Area Network has established guidelines for the ownership of data and other research information. To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed cooperative agreement. Every cooperative or interagency agreement or contract involving the SFAN must include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

> "As the performing organization of this agreement, *institution or organization name* shall follow the procedures and policies set forth in OMB Circular A-110."

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

Researchers should also provide a schedule of deliverables that includes sufficient time for NPS review of draft deliverables before scheduled final submissions.

10. Data Distribution

One of the most important goals of the I&M Program is to integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.

To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the Program. The SFAN will make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will be released with a disclaimer stating as such
- Distributed data are accompanied by complete metadata that clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate distribution
- A complete record of data distribution/dissemination is maintained

To accomplish this, the SFAN will use a number of distribution methods that will allow information collected and developed as part of the program become widely available to park staff and the public.

10.1. Data Distribution Mechanisms

The SFAN's main mechanism for distribution of the Network's inventory and monitoring data will be the Internet. Use of the Internet will allow the dissemination of data and information to reach a broad community of users. As part of the NPS I&M Program, web-based applications and repositories have been developed to store a variety of park natural resource information. The network will use the following applications and repositories to distribute data developed by the program:

- **NatureBib**–a master web-based database housing natural resource bibliographic data for I&M Program parks (NatureBib Home Page)

- **NPSpecies**–a master web-based database to store, manage and disseminate scientific information on the biodiversity of all organisms in all National Park units (NPSpecies Home Page)

- **Biodiversity Data Store**–a digital archive of document, GIS dataset and non-GIS dataset files that document the presence/absence, distribution and/or abundance of any taxa in National Park Service units (Biodiversity Service Center Home Page)

- **Natural Resource and GIS Data Store-**online repository for metadata and associated data products. (**Note:** Currently under development).

- **San Francisco Area Network Website**–provides detailed information about the network and it's I&M Program. Metadata on all inventory and monitoring products developed as part of the Network's I&M plan will be posted to this site. Data and products will either be available through the site, or users will be directed to where the data are stored. (San Francisco Area Network Home Page)

Table 10.1. Data types that will be uploaded to I&M web applications and websites.

Web Application Name	Data types available at site
NPSpecies	Data on Park Biodiversity (species information)
NatureBib	Park Related Scientific Citations
Biodiversity Data Store	The raw or manipulated data and products associated with Inventory and Monitoring data that have been entered into NPSpecies.
NR-GIS Data Store	Metadata and data sets (spatial and non-spatial and products)
SFAN Website	Reports and metadata for all Inventory and Monitoring Data produced by the network.

Currently, the NR-GIS Data Store and Biodiversity Data Store are under development. Until procedures and further guidance become available for the use of these two repositories, the SFAN will disseminate all data developed as part of its I&M Program via the Network website. When both repositories are completely operational, the Network will upload all applicable data and information to each of those sites as needed.

Because network data will reside in the repositories listed above, this data will automatically be searchable via the integrated metadata and image management system and search gateway called NPS Focus. This system is being built with Blue Angel Enterprise software for metadata management and the LizardTech Express Server for image management. Currently ten NPS and two non-NPS databases have been integrated into the NPS Focus prototype in either full or testbed form for one stop searching. NPS Focus has been released as an Intranet version only (http://focus.nps.gov/), a public version is projected in the near future.

Of the ten databases uploaded to date to NPS Focus, NatureBib and the NR-GIS Metadata and Data Store are most applicable to existing network bibliographic and spatial metadata. The network will continue to upload data and information to these two sites, which will coincide with the ability to search for these records through the NPS Focus portal. As NPS Focus reaches further development, other databases and repositories utilized by the network are expected to be searchable through this portal as well.

10.2. Protected Data vs. Public Data

All data and associated information from I&M activities must be assessed to determine their sensitivity. This includes but is not limited to reports, metadata, raw and manipulated spatial and non-spatial data, and maps. Network and park staff must carefully identify and manage any information that is considered sensitive. The Network must clearly identify and define those data needing access restrictions and those to make public.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records, however, are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at http://www.doi.gov/foia/ for further information.

In some cases, public access to data can be restricted. The National Park Service is directed to protect information about the nature and location of sensitive park resources under the following orders and laws:

- Executive Order No. 13007: Indian Sacred Sites, Director's Order #66
- National Parks Omnibus Management Act of 1998 (16 U.S.C. 5937)
- National Historic Preservation Act (16 U.S.C. 470w-3)
- Federal Cave Resources Protection Act (16 U.S.C. 4304)
- Archaeological Resources Protection Act (16 U.S.C. 470hh)

Through these regulations, information that could result in harm to natural resources, including endangered or threatened species, can be classified as 'protected' or 'sensitive' and may be withheld from public release.

The following guidance may be used to determine whether information should be protected:

- Has harm, theft, or destruction occurred to a similar resource on federal, state, or private lands?
- Has harm, theft, or destruction occurred to other types of resources of similar commercial value, cultural importance, rarity, or threatened or endangered status on federal, state, or private lands?
- Is information about locations of the park resource in the park specific enough so that the park resource is likely to be found at these locations at predictable times now or in the future?

- Would information about the nature of the park resource that is otherwise not of concern permit determining locations of the resource if the information were available in conjunction with other specific types or classes of information?
- Even where relatively out-dated, is there information that would reveal locations or characteristics of the park resource such that the information could be used to find the park resource as it exists now or is likely to exist in the future?
- Does NPS have the capacity to protect the park resource if the public knows its specific location?

Resource information that is sensitive or protected requires the:

- Identification of potentially sensitive resources
- Compilation of all records relating to those resources
- Determination of what data must not be released to the public
- Management and archival of those records to avoid their unintentional release

Classification of sensitive I&M data will be the responsibility of the San Francisco Area Network staff, the park superintendents, and investigators working on individual projects. Network staff will classify sensitive data on a case by case, project by project, basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

The Network staff is also responsible for identifying all potentially sensitive resources to principal investigator(s) working on each project. The investigators, whether Network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not release any information in a public forum before consulting with Network staff to ensure that the information is not classified as sensitive or protected.

For example, information may be withheld regarding the nature and/or specific locations of the following resources recognized as ‘sensitive’ by the National Park Service:

- Endangered threatened, rare, or commercially valuable NPS species and habitats
- Minerals or paleontological objects
- Significant caves
- Objects of cultural patrimony

Note that information already in the public domain can, in general, be released to the public domain. For example, the media has reported in detail the return of condors to the Grand Canyon. If an individual requests site-specific information about where the condors have been seen, this information, in general, can be released. However, the locations of specific nest sites cannot be released.

10.3. Access Restrictions on Sensitive Data

The SFAN staff is responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators working on Network projects will be informed that:

- All data and associated information must be made available for review by Network staff prior to release in any format

- Any information classified as protected should not be released in any format except as approved in advance by the National Park Service

When preparing or uploading information into any Network database, the Network staff ensures that all protected information is properly identified and marked. The Network staff works together to ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories are maintained on both secure and public servers, and all records that are marked 'sensitive' during uploading will only become available on the secure servers. Procedures for assigning a sensitivity level to specific records when uploading to both the NPSpecies and NatureBib databases are discussed in the NPSpecies and the NatureBib Data Management Plans.

Thus, access to data on sensitive park resources can be limited to Network staff or research partners. However, limits to how these data are subsequently released must also be clearly defined. It is crucial that the Network staff institute quality control and quality assurance measures to ensure that the person doing the uploading of records into the online applications is familiar with the procedures for identifying and entering protected information.

10.3.1. Public Access to Network Inventory and Monitoring Data

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the internet. Network project data will be available to the public at one or more internet locations:

- The San Francisco Bay Area Network web site
- Public servers for the NPSpecies and NatureBib databases
- Public server for the Biodiversity Data Store
- Public server for the NR-GIS Data Store

The Network will regularly provide updated information about inventories and monitoring projects, including annual reports and detailed project reports through the Network web site. Information on species in the National Parks, including all records generated through the Inventory and Monitoring Program, will be maintained and assessable through the NPSpecies

database. Bibliographic references that refer to National Park System natural resources will be accessible through the NatureBib database. Documents, maps, and data sets containing resource information from all sources, and their associated metadata, will be accessible through the Biodiversity Data Store and/or NR-GIS Data Store. Each of these databases/repositories will be accessible via both a secure server and a public server, and the public can access all information in these databases except those records marked as 'sensitive.'

10.4. Data Availability

Both raw and derived data resulting from the Network's inventory and monitoring projects will be fully documented with FGDC compliant metadata and made available to the public via the Network's website. The metadata for all datasets will be made available to the public as soon as they are provided and verified by the Investigator(s) or project managers.

Datasets for short-term studies (inventories) will be provided to the public on the SFAN website, two years following the year the data were collected or following publication of the Investigator's results (whichever comes first). Long-term (monitoring) studies will be provided to the public in four or five year intervals, or when trend analyses have been completed and reported on by the network. (This will be specific to each network monitoring protocol, refer to the Network's implementation plan for further information). Before data are posted, the Investigator or project leader will be asked to verify the final dataset and metadata if necessary. Once the Network staff and Investigator verify the data set, the data will be made accessible to the public, provided no sensitive information is identified.

SFAN staff will notify investigators prior to making data sets available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the dataset by the public. Network staff will review the Investigator's request and determine whether the request will be granted and for how long the dataset will remain restricted.

10.4.1. Data Acquisition Policy

The SFAN will develop a dataset acquisition policy that will be made available to all SFAN website users who wish to acquire program data and information. This policy will include such things as:

- A mandatory web-based questionnaire to be completed before acquiring data (This questionnaire will allow Network staff the ability to maintain a distribution log specifying recipient name and contact information, intended use of data, export file format, delivery date and method, and data content description noting range by date and geography of data delivered.)

- A statement about use and appropriate citation of data in resulting publications

- Request that acknowledgement be given to the National Park Service Inventory and Monitoring Program within all resulting reports and publications

All data sets with public access available on the SFAN website will be accompanied by the Network's acquisition policy.

10.5. Data Feedback Mechanisms

The SFAN website will provide an opportunity for NPS staff, cooperators and the public to provide feedback on data and information gathered as part of the Network's I&M Program. A "comments and questions" link will be provided on the main page of the site for general questions and comments about the Network's program and projects. A more specific "data error feedback" link will direct comments to the SFAN staff pertaining to errors found in website accessible data. Annual reporting of progress will be presented to the Board of Directors and to the Technical Committee on a yearly basis, and feedback will be expected during and following these presentations.

10.5.1. Data Error Feedback Procedures

The following feedback procedures describe the process which the SFAN will use to receive and verify data errors identified by public and private data users:

- Web Users send in a notification about an alleged error through the SFAN website. SFAN staff then sends an acknowledgment to the notifier.

- SFAN staff then inputs the information into a data error log table incorporated in either each of the network monitoring databases or a specific error tracking database developed for the network.

- SFAN then determines if the data questioned by the notifier are correct or incorrect. If the data are correct, then the Network staff informs the notifier that no corrections are to be made and the information stands. If the data are incorrect, the Network staff makes the appropriate corrections and notifies the original data collectors (cooperator, other agency, park staff, etc…).

- Once data are corrected, the Network website will be refreshed with the corrected information.

- Throughout this process, the Network staff will continually inform the notifier via e-mail of the status.

11. Data Maintenance, Storage and Archiving

This chapter describes procedures for the long-term management and maintenance of digital data, documents, and objects that result from SFAN projects and activities. The overall goals of these procedures are:

- to avert the loss of information over time
- to ensure that our information is properly interpreted by a broad range of users
- to ensure that our information can be easily obtained and shared through future decades

Direction for managing these materials (as well as digital materials) is provided in NPS Director's Order 19: Records Management (2001) and its appendix, NPS Records Disposition Schedule (NPS-19 Appendix B, revised 5-2003). NPS-19 states that all records of natural and cultural resources and their management are considered mission-critical records; that is, necessary for fulfillment of the NPS mission. NPS-19 further states:

> Mission critical records are permanent records that will eventually become archival records. They should receive the highest priority in records management activities and resources and should receive archival care as soon as practical in the life of the record.

Section N of NPS-19 Appendix B, which provides guidelines on natural resource-related records (including, specifically, the results of Inventory and Monitoring Programs), indicates that all natural resource records are considered "permanent," that is, are to be transferred to the National Archives when 30 years old. It also indicates that non-archival copies of natural resource-related materials are "...potentially important for the ongoing management of NPS resources" and should not, in any instance, be destroyed.

Effective long-term data maintenance is inseparable from proper data documentation, and an essential part of any archive is accompanying explanatory materials (Olson and McCord 1998). This chapter will refer to, and in some cases elaborate on, metadata standards and dataset documentation procedures that are more fully explained in Chapter 7 of this plan.

11.1. Digital Data Maintenance

In general, digital data maintained over the long term will be one of two types: short-term data sets, for which data collection and modification have been completed (i.e., inventory projects); and long-term monitoring data sets, for which data acquisition and entry will continue indefinitely.

Following the lead of the National Park Service and the national I&M program, the SFAN has adopted MS-Access as its database standard and ArcGIS as its spatial data management standard.

The SFAN will remain current and compatible with the NPS or national I&M version standards for these software programs.

Technological obsolescence is a significant cause of information loss, and data can quickly become inaccessible to users if stored in out-of-date software programs or on outmoded media. Maintaining digital files involves managing the ever-changing associated infrastructure of hardware, software, file formats, and storage media. Major changes in hardware can be expected every 1-2 years, and in software every 1-5 years (Vogt-O'Connor 2000). As software and hardware evolve, data sets must be consistently migrated to new platforms.

11.1.1. Short-term data sets

For short-term data sets created or managed by the SFAN, upon project finalization the native version of the dataset (typically in database or spreadsheet format) will be archived in read-only format with accompanying metadata, other data documentation, protocols, and final reports. These file(s) will be accompanied by a readme.txt file that explains the contents of each file and file relationships. All finalized files will be stored both in the park server archives and in the SFAN directory archive section in the appropriate project folder.

The SFAN will also update completed and archived data sets that may be in older versions of MS-Access, with the goal of having no dataset more than two versions behind the current version used by the SFAN. There is the risk of losing a certain amount of performance in the process of conversion; for example, complex data entry forms or reports may not function properly in an upgraded version. To the extent possible, proper functionality of data entry forms and reports will be maintained; however, the priority will be to ensure basic table and relationship integrity. All previous versions of the data set will be saved.

11.1.2. Long-term monitoring data sets

Long-term monitoring data sets require regular updates and conversion to current database formats. All active or long-term databases will conform to the current NPS and I&M software version standards.

Long-term monitoring projects will also have annual data archiving requirements. Project managers will forward verified and validated data sets to the Network Coordinator according to schedules detailed in the data management SOPs for each monitoring project. Archived long-term data sets or subsets will be saved in their native formats, with appropriate metadata, as discussed for short-term data sets.

11.1.3. Version control

Previous versions of databases will be saved in their native format and archived in addition to the current version. Documentation of version updates and associated details will be part of the archive metadata document. Whenever possible, revision information and history should be included in tables within the database files themselves as well. File names of the archived revisions should clearly indication the revision number or date.

11.1.4. Spatial data

Spatial data sets that are essential to the SFAN will be maintained in a format that remains fully-accessible by the current ArcGIS version. ArcGIS has maintained compatibility with previous data formats, and while shapefiles have retained all functionality in ArcGIS, coverages may require conversion to ArcGIS format if they are no longer supported. At this time there is no practical way to save GIS data in a software or platform-independent format.

11.2. Storage and Archiving Electronic Data and Documents

SFAN park managers will require access to digital archives generated by the I&M program in order to make informed natural resource management decisions, review program achievements, and assess ecological trends of park resources. Because access to the I&M archives is limited to direct Network personnel, separate digital archives are required on the park LANs.

Digital archives of completed I&M products, including SOPs, reports, and data sets, will be maintained at the park and Network levels. For long-term monitoring projects, data sets will be uploaded to both digital archive locations on an annual basis according to schedules outlined in the SOPs.

11.2.1. Network digital archives

Only I&M staff and system administrators have permission to access files on the I&M server, and restrictions have been established on archived data files. Directories containing completed project data or interim versions of ongoing projects are designated as read-only for all staff with the exception of the lead data manager. In this way, any changes must be routed through the lead data manager, who is responsible for ensuring that documentation and readme files associated with the data set are updated.

11.2.2. Electronic archive procedures

Final digital products, including SOPs, final reports, and data sets will be placed in a read-only format in the SFAN digital archives located on the Headlands server, Network I&M directory. The SFAN digital archives back-up schedule follows that outlined in Section 4.1.4 for server back-up specifications and schedules.

Only final documents will be archived – no drafts or works in progress. Microsoft Word documents will be converted to PDF files prior to archiving. Only the lead data manager will archive data sets. Data sets must be validated and verified, must represent a complete set of records, and must have accompanying readme text files. Once final data and reports have been submitted, interim or draft products do not need to be maintained.

Archived digital files will then be entered into a local desktop version of NatureBib, which is periodically uploaded into the national online version of this database by the lead data manager.

Even though NatureBib is a bibliographic database designed to keep track of documents, it does provide an interface for entering information on datasets as well. As such, it enables users to interact with only one interface no matter what kind of information (a document or dataset) is to be entered and thus saves the user from having to enter information into two both NatureBib and Dataset Catalog. The lead data manager is currently investigating the potential to link the desktop version of NatureBib with Dataset Catalog to avoid duplicate data entry into the two databases.

The desktop version of NatureBib will include a hotlink field that directs users to documents and data sets located in the SFAN digital archives.

To ensure long term management of and protection for the work that is generated by the I&M program, a hardcopy of final products will be sent to the Records Center located in the Presidio. Final versions of documents are labeled by the project manager and placed in a drop box in Building 1061 at Fort Cronkite so that they can be transported, on a regular basis, to the Records Center.

In order to determine whether a document or data set merits archiving, the SFAN staff will follow guidelines presented in Director's Order #19: Records Management and a NatureBib flowchart found at http://www.nature.nps.gov/nrbib/flowchart.doc.

11.2.3. Directory structure for archived project folders

The current directory structure for archived SFAN data files, including GIS data, is mirrored by the Network directory structure presented in Appendix B. The organization and naming of archived I&M folders and files will be intuitive to users unfamiliar with a specific project. Because each project will have its own variations and idiosyncrasies, a standardized structure isn't realistic; however, all completed project archives will include most of the following elements:

- administrative documents such as agreements, contracts, correspondence, research permits, etc.
- final data; native format
- final report
- project metadata (including a Dataset Catalog report)
- readme files – includes an explanation of directory contents and information on updates or changes.

Ongoing monitoring projects will include several categories in addition to the above list:

- protocol documents
- SOPs
- annual reports
- conceptual or statistical models used for data interpretation
- data "snapshots" and/or raw data files used as source data
- project metadata (including a Dataset Catalog report)
- readme files – includes an explanation of directory contents and information on updates or changes.

11.3. Storage and Archiving Hardcopy Documents and Objects

This section applies to documents such as final reports, program administrative documents, contracts and agreements, and other documents related to the SFAN I&M program for which digital copies are unavailable. This section also applies to physical items such as natural history specimens, photographs, or audio tapes. In most instances these documents and objects are essential companions to the digital archives described above.

Curators for parks within the SFAN are an ongoing source of expertise, advice, and guidance on archiving and curatorial issues, and they have a role in almost all projects undertaken by the network. Project managers should involve park curators when projects are in the planning stage to ensure that all aspects of specimen curation or document archiving is considered, and that any associated expenses are included in project budgets.

Because the SFAN offices are located at the GGNRA, archive procedures for SFAN hardcopy documents and objects comply with those established and followed by the GGNRA.

11.3.1. Documents

Beginning in October 2004, documents that are not available in digital format will first be scanned and saved as PDF files in the SFAN Archive directory. The process for archiving these documents will then follow that outlined in Section 11.2 for electronic digital archives, which includes entering the document into NatureBib and forwarding the original document to the GOGA Records Center.

At this point, no effort will be made to scan hard-copy only documents received by or produced by the SFAN prior to October 2004. Scanning to digital format will only occur as needed. In order that hard-copy documents are readily available to park staff, a local library is instead being developed in Building 1061 at Fort Cronkite. The library will house copies of relevant documents filed according to the author's last name. Original copies will forwarded to the GOGA Records Center and records will be entered into NatureBib, as described previously.

11.3.2. Specimens

Specimens collected under the auspices of SFAN inventory and monitoring projects will be cataloged and maintained according to Director's Order #24: NPS Museum Collections Management. Specific repositories for specimens are detailed in the inventory contracts and study plans and monitoring SOPs.

11.3.3. Photographs

Several of the SFAN I&M projects incorporate digital and/or print photography into their protocols. Photo documentation may be an integral part of an I&M protocol which relies heavily on motion-sensor cameras, as is the case with a multi-species vertebrate inventory . In some

cases, however, photography is supplemental, used to document unusual events or rare species encountered.

Archiving procedures for digital and print photos will follow guidelines previously established for digital and hard-copy documents, respectively.

12. References

Brunt, J.W. 2000. Data management principles, implementation and administration. Pp. 25-47 *in* Michener, William K., and James W. Brunt (eds), 2000, Ecological Data: Design, Management and Processing, Methods in Ecology Series, Malden, MA: Blackwell Science Ltd.

Boetsch, J.R., B. Christoe, and R.E. Holmes. 2004. Draft data management plan for the North Coast and Cascades Network Inventory and Monitoring Program. USDI National Park Service. Port Angeles, WA. 89 pp.

Chapal, Scott E., and Don Edwards. 1994. Automated smoothing techniques for visualization and quality control of long-term environmental data. Pp. 141-158 *in* Michener, William K., James W. Brunt, and Susan G. Stafford (eds), 1994, Environmental Information Management and Analysis: Ecosystem to Global Scales. London: Taylor & Francis Ltd.

Executive Order No. 12906: Coordinating Geographic Data Acquisition and Access. April 11, 1994.

Executive Order No. 13007: Indian Sacred Sites. May 24, 1996.

Federal Geographic Data Committee. (1998). FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C.

Federal Geographic Data Committee. (2000). Content Standard for Digital Geospatial Metadata Workbook Version 2.0. Federal Geographic Data Committee. Washington, D.C.

Michener, W.K., and J.W. Brunt. 2000. Ecological Data: Design, Management and Processing, Methods in Ecology Series, Malden, MA: Blackwell Science Ltd.

Miller, A.B. 2001. Managing data to bridge boundaries. Pp. 316-320 *in* Harmon, David (ed), Crossing Boundaries in Park Management: Proceedings of the 11th Conference on Research and Resource Management in Parks and on Public Lands, Hancock, MI: The George Wright Society.

National Park Service. In Progress. Director's Order #66: FOIA and Protected Resource Information.

National Park Service. 1998. National Parks Omnibus Management Act of 1998 (PL 105-391): National Park Service Concessions Management Improvement Act of 1998.

National Park Service. 2000. Director's Order #24: Museum Collections Management.

National Park Service. 2001. Director's Order #19: Records Management.

National Park Service. 2002. Director's Order #11B: Ensuring quality of information disseminated by the National Park Service.

National Park Service. 2003. NPS records disposition schedule. NPS-19, Appendix B – Revised, 5-03. 61 p.

Olson, R.J. and R.A. McCord. 1998. Data Archival. In: Michener, W.K., J.H. Porter, and S.G. Stafford. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque, NM. pp 53-57.

Palmer, Craig J. 2003. Approaches to quality assurance and information management for regional ecological monitoring programs. Pp. 211-225 *in* Busch, David E., and Joel C. Trexler (eds), 2003, Monitoring Ecosystems: Interdisciplinary Approaches for Evaluating Ecoregional Initiatives. Washington, DC: Island Press.

Palmer, Craig J., and Eric. B. Landis. 2002 Draft. Lake Mead National Recreational Area Resource Management Division: Quality System Management Plan for Environmental Data Collection Projects.

Tessler, Steve, and Joe Gregson. 1997. Draft Data Management Protocol. NPS I&M Program. http://www1.nrintra.nps.gov/im/dmproto/joe40001.htm

Vogt-O'Connor, D. 2000. Planning digital projects for preservation and access. National Park Service Conserve O Gram 14:4. NPS Museum Management Program, Washington, DC. 4 p.

U.S. Environmental Protection Agency. 2003. Guidance for geospatial data quality assurance project plans. EPA QA/G-5G. EPA/240/R-03/003. Washington, DC: Office of Environmental Information.

Appendix A. Data Stewardship Framework

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
Project Technician	Collect, record, and verify data.	Obtain training in data management for the project. Read and follow project protocols, study plans, and relevant NPS guidance. Communicate with Project Manager and Data Manager. Record measurements and observations based on project objectives. Enter and verify observed or measured data values in project database. Schedule and perform regular data transfer and backup. Assist with data and procedural documentation, especially deviations from the protocol or study plan.	Aware of specific protocol and related subject(s) for the project. Aware of related protocols and projects. Aware of Network activities, NPS I&M Program generally.
Project Manager	Oversee and direct project operations.	Communicate with Project Technicians, Data Manager, and I&M Network Coordinator. Complete project documentation describing the who, what, where, when, why and how of a project. Develop, document and implement standard procedures for field data collection and data handling. Enact and supervise quality assurance and quality control measures for the project. Supervise and certify all field operations, including staff training, equipment calibration, species identification, and data collection. Supervise or perform data entry, verification and validation. Maintain concise explanatory documentation of all deviations from standard	Aware of specific protocol and related subject(s) for the project. Aware of related protocols and projects. Aware of Network activities, NPS I&M Program generally. Aware of park resource information management and GIS Plans and capabilities. Aware of other related projects external to NPS.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
		procedures. Ensure documentation of important details of each field data collection period. Maintain hard copies of data forms and send original data forms to archive on a regular basis. Work with program coordinators to identify analysis and reporting mechanisms, and to establish a schedule for regular project milestones such as data collection periods, data processing target dates, and reporting deadlines. Produce regular summary reports and conduct periodic trend analysis of data, store the resulting reports, and make them available to users. Act as the main point of contact concerning data content.	
Project Manager (Continued)		**The project manager works closely with the data manager to:** Develop quality assurance and quality control procedures specific to project operations. Identify training needs for staff related to data management philosophy, database software use, quality control procedures, etc. Coordinate changes to the field data forms and the user interface for the project database. Generate complete data documentation (ie. metatdata) and maintain master data. Identify sensitive information that requires special consideration prior to distribution. Manage the archival process to ensure regular archival of project documentation, original field data, databases, reports and summaries, and other products from the project.	

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
		Define how project data will be transformed from raw data into meaningful information, and to create data summary procedures to automate and standardize this process. Identify and prioritize legacy data for conversion to a modern format. Increase the interpretability and accessibility of existing natural resource information.	
Park Data Manager	Ensure inventory and monitoring data is organized, useful, compliant, safe, and available.	Provide overall Network planning, training, and operational support for the awareness, coordination, integration of data and information management activities, including people, information needs, data, software, and hardware. Review and approve all data acquisition plans, hardcopy and electronic field forms, and data dictionaries. Communicate with Project Manager, I&M Network Coordinator, and GIS specialist. Develop and maintain overall Network and individual Vital Sign data management operating guidelines and relationship to national standards and procedures. Develop and maintain the infrastructure for metadata creation, project documentation, and project data management. Create and maintain project databases in accordance with best practices and current program standards. Provide training in the practice of data management, tailored to the needs of project personnel. Develop ways to improve the accessibility of digital data. Establish and implement procedures to protect sensitive data according to	Aware of specific protocol and related subject(s) for all network projects. Aware of related protocols and projects. Aware of park resource information management and GIS Plans and capabilities. Aware of other related projects external to NPS.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
		project needs. Establish and implement procedures to convert legacy data sets to modern formats. Review Network databases and data sets for data standards. Collaborate with GIS specialists to integrate tabular data with geospatial data in a GIS system in a manner that meets project objectives. Install, maintain, and support specific database software applications and NPS database applications. Work with information technology specialists to resolve hardware and software issues.	
Park Data Manager (Continued)		**Data managers will also work closely with the project manager to:** Define the scope of the project data, and create a data structure that meets project needs. Become familiar with how the data is collected, handled and used. Develop quality control and quality assurance aspects of project protocols. Develop data management sections, including database documentation, of formal project Standard Operating Procedures. Identify elements that can be built into the database structure to facilitate quality control, such as required fields, range limits, pick lists and conditional validation rules. Create a user interface that streamlines the process of data entry, review, validation, and summarization, and that is consistent with the capabilities of the project staff.	

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
		Develop automated database procedures to improve the efficiency of the data summarization and reporting process. Make sure that project documentation is complete, complies with metadata requirements, and enhances the interpretability and longevity of the project data. Ensure regular archival of project materials. Inform project staff of changes and advances in data management practices.	
Lead Data Manager	Ensure inventory and monitoring data is organized, useful, compliant, safe, and available.	Perform all regular duties of park data manager. Serve as Point of Contact for National Park Service database applications (NPSpecies and NatureBib). Communicate with national-level I&M Program for updates on NPS database applications and data standards. Initiate and lead regular SFAN data management workgroup meetings. Assist Network Coordinator with maintaining Network digital archives and with annual reporting requirements.	Aware of specific protocol and related subject(s) for all network projects. Aware of related protocols and projects. Aware of park resource information management and GIS Plans and capabilities. Aware of other related projects external to NPS.
GIS Specialist	Support park management objectives with GIS and resource information management.	Coordinate and integrate local GIS and resource information management with Network, regional, and National standards and guidelines. Perform required level of technical data management and/or GIS activities, including data conversion and data documentation.	Aware of Network activities, NPS I&M Program generally. Aware of data management relationships between parks and the network. Aware of associations between park resource issues

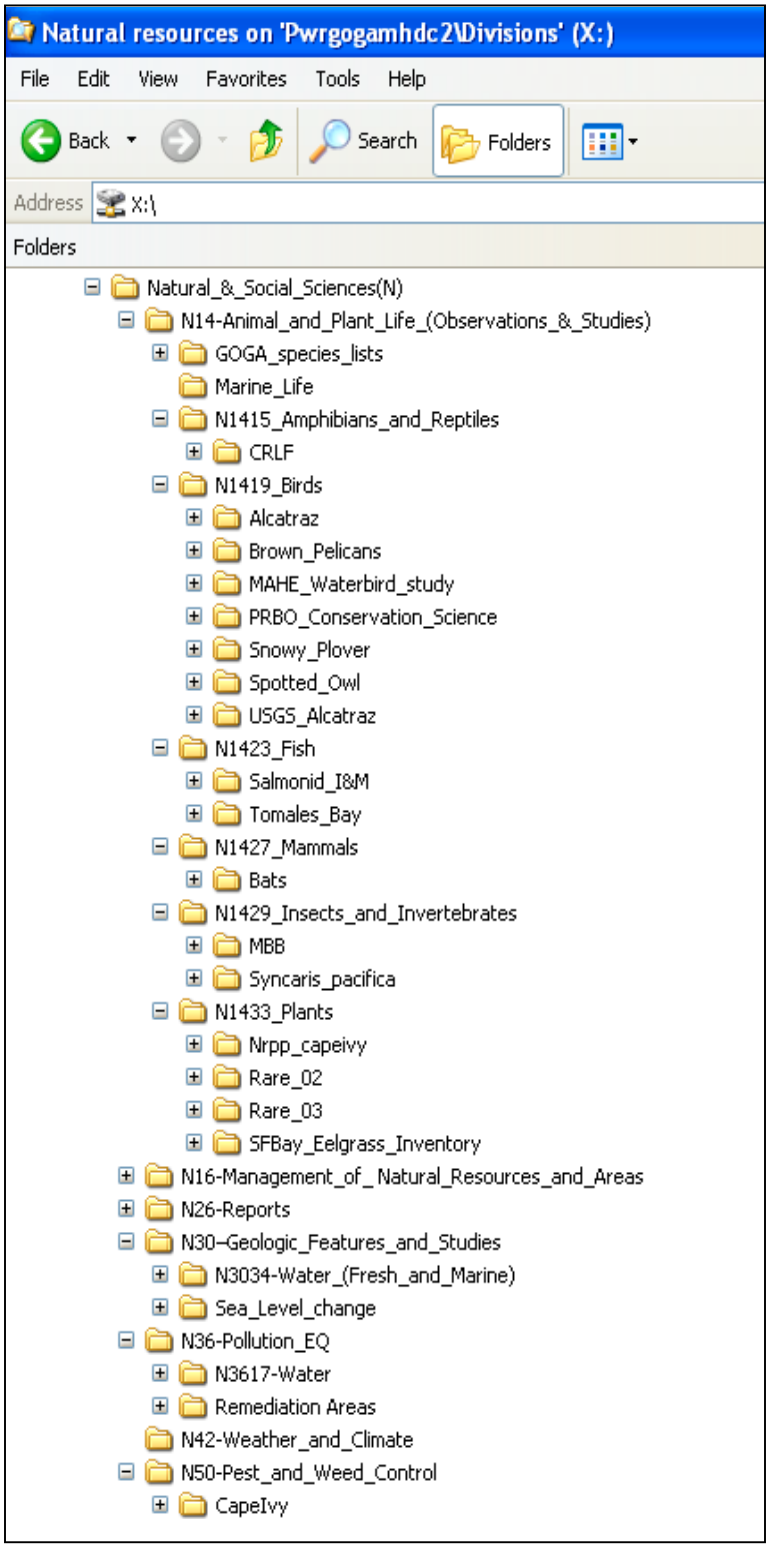
Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
			and park/network I&M objectives.
GIS Specialist (continued)		**The GIS specialists will work in collaboration with project managers to:** Determine the GIS data and analysis needs for the project. Develop procedures for field collection of spatial data including the use of GPS and other spatial data collection techniques. Display, analyze, and create maps from spatial data to meet project objectives. Properly document data in compliance with spatial metadata standards. **GIS specialists will also work directly with data managers to:** Design databases and other applications for the network. Create relationships between GIS and non-spatial data, and create database and GIS applications to facilitate the integration and analysis of both spatial and non-spatial data. Establish and implement procedures to protect sensitive spatial data according to project needs. Develop and maintain an infrastructure for metadata creation and maintenance. Ensure that project metadata are created and comply with national and agency standards.	
Network Ecologist	Integrate science in network activities	Ensure useful data is collected and managed by integrating natural resource science in network activities and products, including objective setting, sample design, data analysis, synthesis, and reporting. Work with the network data manager to incorporate data management in	Aware of park resource information management and GIS Plans and capabilities.

Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
		monitoring protocols. Guides and/or performs statistical and other analyses of network data. Contributes to the synthesis and reporting of data and information.	Aware of database tools and applications used by the I&M program. Aware of Data Stewardship principles.
Network Coordinator	Coordinate all network activities	Ensure programmatic data and information management requirements are met as part of overall Network business. Communicate with Network staff, park staff at all levels, and other appropriate audiences to support and emphasize data management as a critical aspect of network business. Work with lead data manager regarding data management policy and guidelines, budget, staffing, and training. Hold Network staff accountable for responsibilities involving data management.	Aware of database tools and applications used by the I&M program. Aware of data management principles.
Information Technology/ Systems Specialist	Provide IT/IS support.	Provide and maintain an information systems and technology foundation to support data management. Advise project participants about capabilities of hardware and software resources to support project and program objectives. Work with database manager to resolve hardware and software issues relating to database functions and availability.	Aware of NPS I&M Program generally. Aware of database tools and applications used by the I&M program.
I&M Data Manager (National Level)	Provide service-wide database availability and support	Provide service-wide database design, support, and services, including receiving and processing to convert, store, and archive data in service-wide databases. Work with network data manager to resolve local issues involving the access and use of inventory and monitoring databases.	Aware of San Francisco Area Network contacts
Other End Users	Use and apply Network services and products	These 'information consumers' include park managers and superintendents, researchers, staff from other agencies, and the public. End users at all levels are generally responsible for providing necessary and requested feedback, review, and comments on various products in order to	Varies with each end user. Awareness is an outcome of end use.

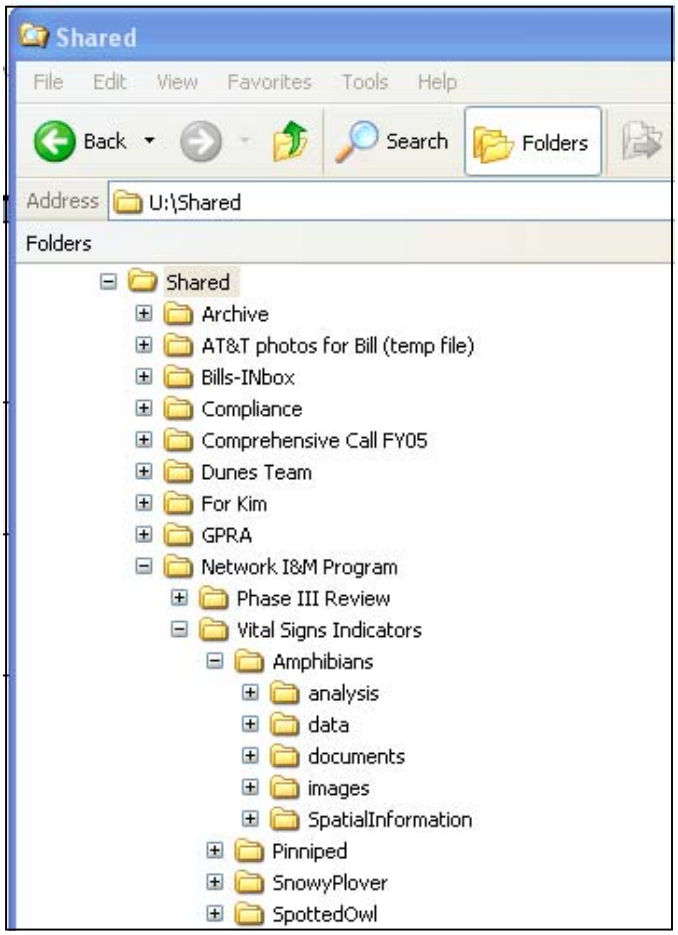
Role	Programmatic Responsibility	Data Stewardship Responsibilities	Recommended Awareness Level
		sustain the continuous improvement of network operations and services.	

Appendix B. Digital File Folder Directory Structures

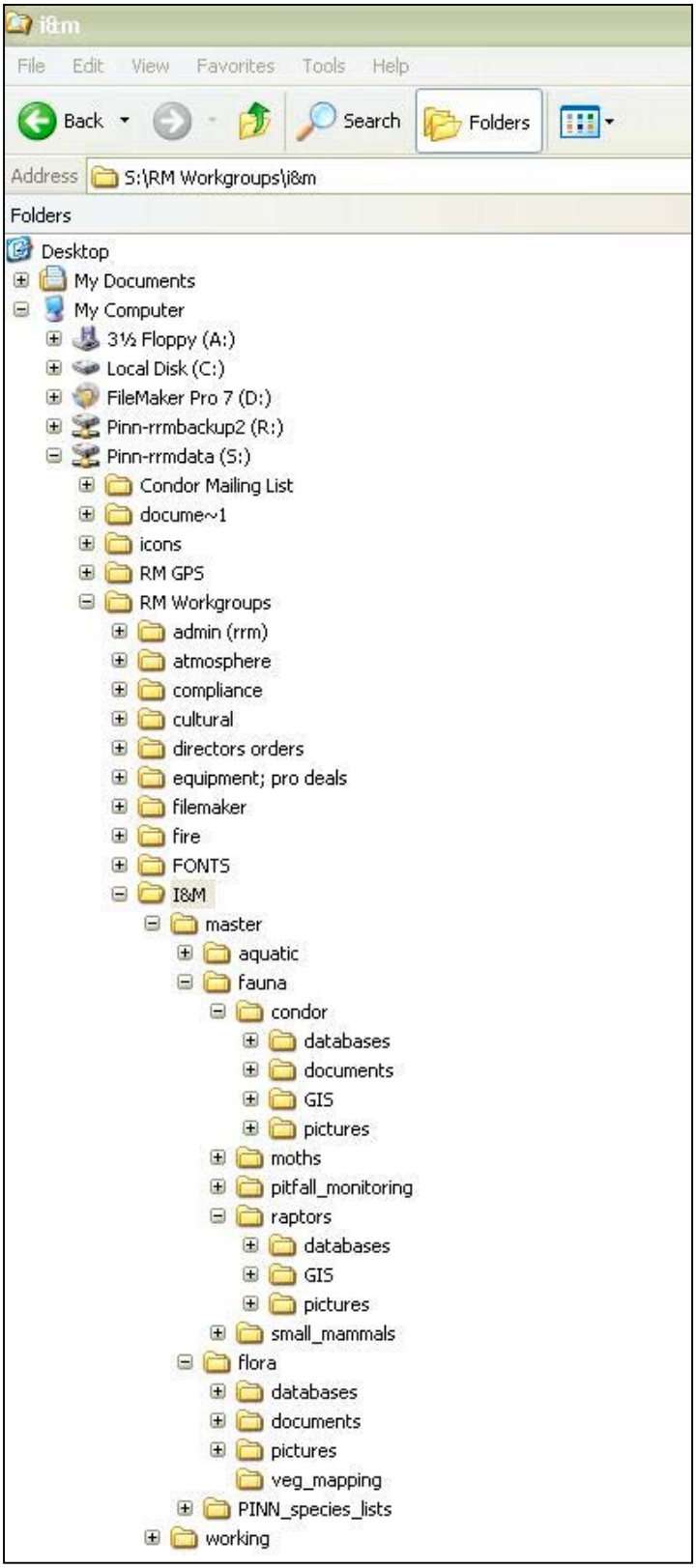
Golden Gate National Recreation Area



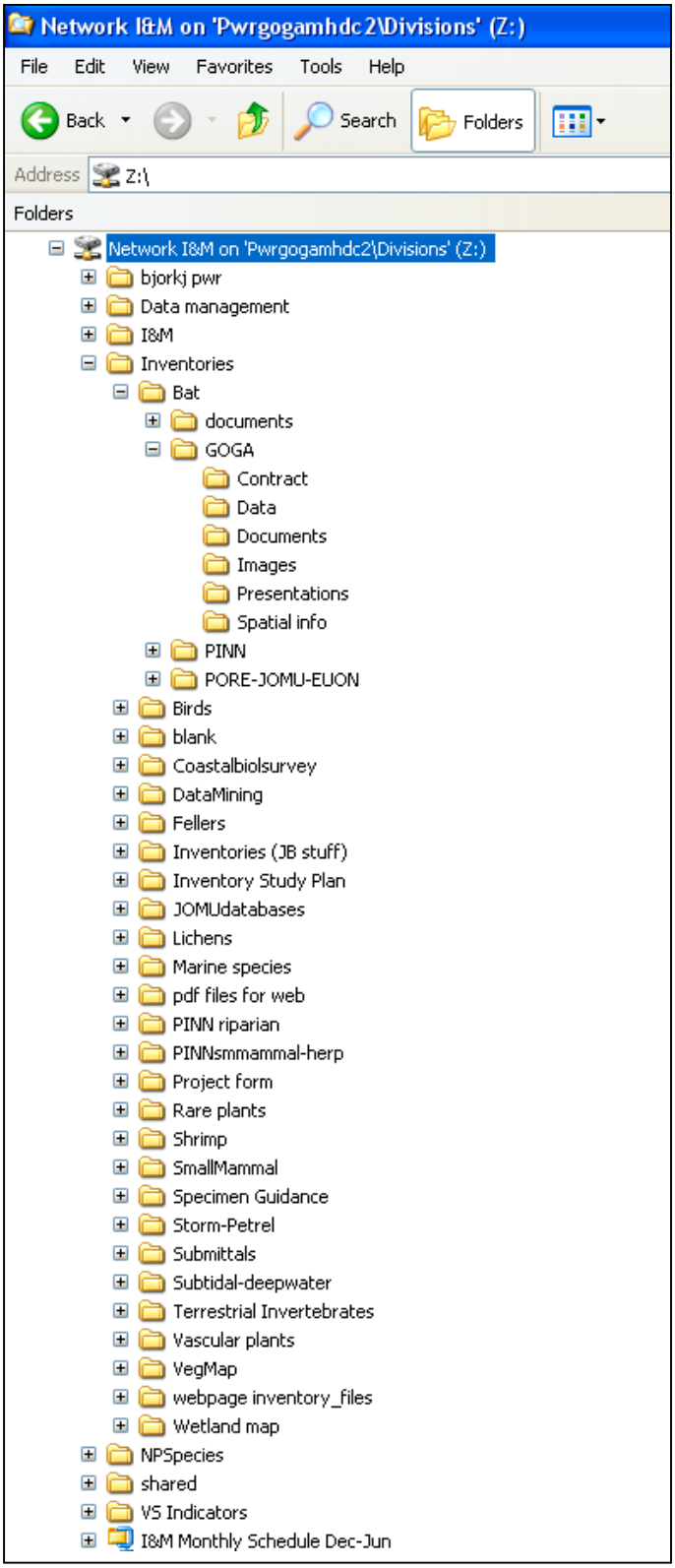
Point Reyes National Seashore



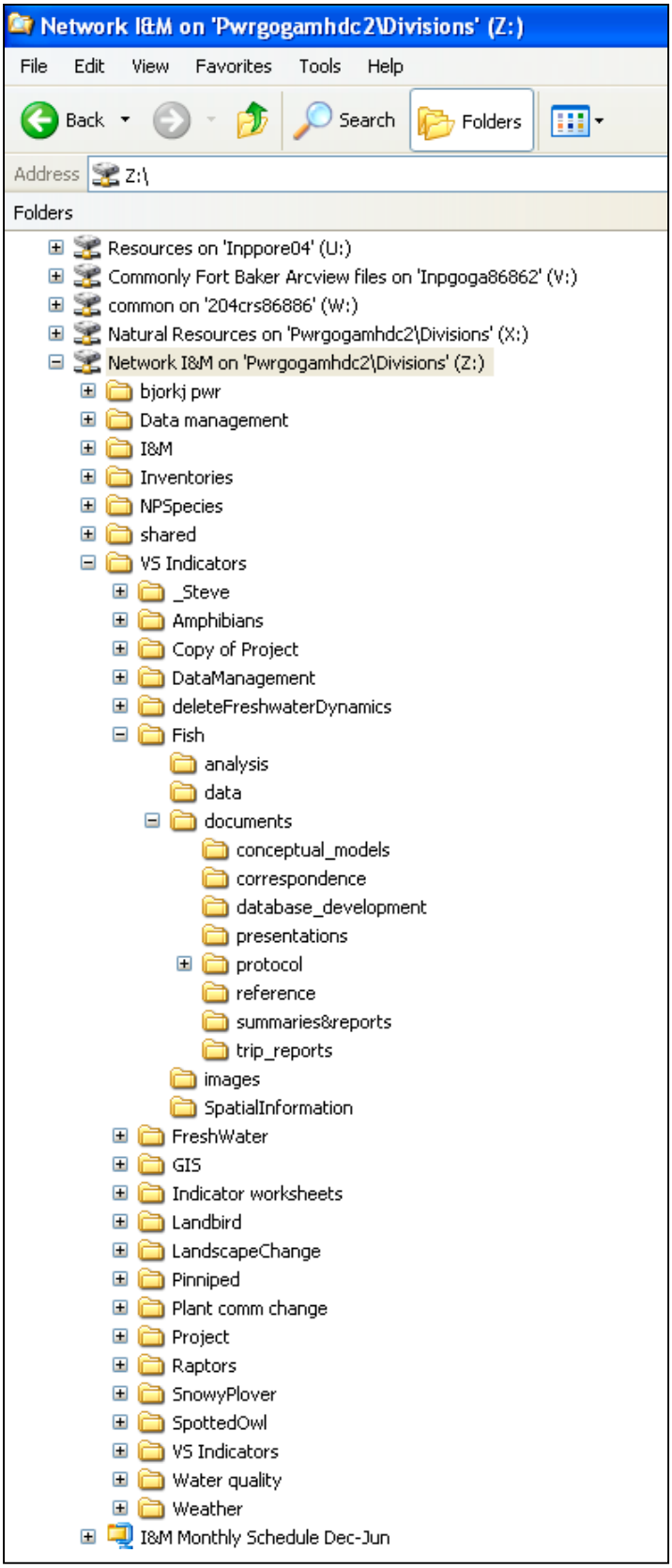
Pinnacles National Monument



SFAN Network I&M Digital File Folder Directory Structure – INVENTORY



SFAN Network I&M Digital File Folder Directory Structure – MONITORING



Appendix C. Database Design Standards

DATABASE INFRASTRUCTURE

Relational database guidelines adopted by the SFAN I&M Program accommodate a diversity of data types, eliminate redundancy, maintain integrity among data tables, and ensure that data can be combined for integrative analysis and reporting. The database system is composed of multiple, separate project databases designed in MS Access with a similar database structure. Project databases are stored on servers at the host park unit. Each project database functions independently, but common database elements allow data to be combined from multiple databases for comprehensive analysis and reporting purposes. The standardized structure of SFAN I&M databases is based on the National Park Service's Natural Resource Database Template.

The standard database design consists of interrelated core tables, shared lookup tables, and project specific lookup tables and field/sampling tables. Core tables are distinguished from common lookup tables in that they reside in each individual project database and are populated locally. Common lookup tables, on the other hand, are stored in a central location such that each project database is linked to and refers to the shared lookup tables. Project specific lookup tables refer to locally stored tables intended to standardize and ensure the quality of data entry. Some project specific lookup tables may be unique to a project database, while others are standardized and are an integral component of several related project databases. Field/sampling tables are stored locally, are unique to the project database, and contain data collected or recorded specific to the goals of the project.

CORE TABLES

Two core tables – Events and Locations – have been designed and standardized across project databases with regard to required fields, field names, and data types. The Locations table describes the site of data collection – its unique identifying code, a description of the site, UTM coordinates, elevation, etc. The Events table generally represents the "header" information from a particular sampling event, site visit, etc. Included in the Events table are fields such as date, start time, observer, and weather – data that is common across all additional, more specific data collected during the sampling event.

COMMON LOOKUP TABLES

Three common lookup tables have been designed to share across all project databases, although each park unit maintains its own set of these tables. The tables are maintained by the park Data Managers and stored on readily accessible servers. The tables are incorporated into project databases as linked tables. The Parks table simply provides informational data on the NPS unit where the data was collected. Similarly, the Project table provides information as to the nature of the monitoring project, such as the name of the project manager. A third table, the Species table, will be generated for use in project databases from the finalized NPSpecies database. The Species table will only be required for project databases for which information is collected on

more than one species. Project managers will be informed when the Species table is available for their use.

Other common lookup tables may be designed and shared between similar project databases. Vegetation related databases, for example, may be linked to a standard cover class lookup table. Such tables are maintained by project managers.

PROJECT SPECIFIC LOOKUP TABLES

One project specific lookup table has been standardized across all project databases. An observers lookup table is employed to ensure a standardized method of entering observer names or codes. Observers listed in the table include park staff and volunteers collecting data for a particular project. In cases where different project databases pull from the same list of observers, a shared observers table can be used. Projects that have a variable number of observers present on any given field day should enter observers in a separate table linked to the Events table by a one-to-many relationship. Projects with static observer numbers may enter values directly into the Events table.

Other project specific lookup tables should be designed for fields that have a limited number of discrete data entry options, such as a vegetation community type or gps unit used. The use of such lookup tables facilitates data entry and serves as a quality control mechanism.

FIELD / SAMPLING TABLES

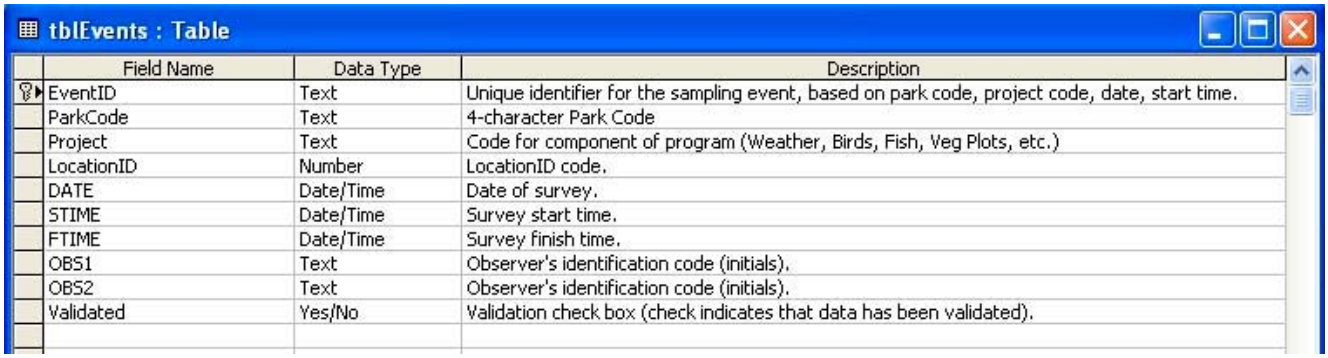
All relevant data collected in the field is entered in one or several field/sampling tables, each linked back to the Events table via the unique EventID field. These tables often contain the actual data you set out into the field to collect. For example, a plant composition table may list the plant species and percent cover observed in a particular vegetation plot. A fish table could list the locations, number, and behavior of fish observed during a particular survey of a stream reach. Data such as these, collected in a repetitive manner during a single sampling event, is linked to the Events table with a one-to-many relationship. Field data collected may also be collected in a one-to-one relationship with the Events table. In a water quality database, for example, water quality parameters are measured only once during a particular, unique site visit. Data of this type is also stored in a separate field/sampling table.

TABLE STRUCTURES

The following table designs should be applied to all SFAN I&M databases. Clarification and important points to consider are listed below each table design.

Core Tables

1. **Events Table (tblEvents)**



tblEvents : Table

Field Name	Data Type	Description
EventID	Text	Unique identifier for the sampling event, based on park code, project code, date, start time.
ParkCode	Text	4-character Park Code
Project	Text	Code for component of program (Weather, Birds, Fish, Veg Plots, etc.)
LocationID	Number	LocationID code.
DATE	Date/Time	Date of survey.
STIME	Date/Time	Survey start time.
FTIME	Date/Time	Survey finish time.
OBS1	Text	Observer's identification code (initials).
OBS2	Text	Observer's identification code (initials).
Validated	Yes/No	Validation check box (check indicates that data has been validated).

EventID: By example, the following text string represents an acceptable EventID, "GOGA-PLOVER-1/10/1996-9:10:00 AM". The EventID field may be automatically populated using Visual Basic code or a macro after the park code, project code, date, and time fields have been entered.

ParkCode: Set default value to your park unit so that manual data entry is not required.

ProjectCode: Set default value to your project code so that manual data entry is not required.

LocationID: Unique identifying code for the location of data collection. Best if created as a lookup field to the locations table.

Date: Set to date/time data type with a "short date" format.

Time: Set all time values to date/time data type with a "short time" format.

Observers: Unique identifying code for the observer or data collector. Best if created as a lookup field to the observers table. More than one observer for a specific sampling event must be separated into separate fields (i.e., Obsv1, Obsv2).

Validated: A check box that indicates that the data contained in the record has been validated at the end of the sampling season. Only validated data will be archived.

2. Locations Table (tblLocations)

tblLocations : Table

Field Name	Data Type	Description
LocationID	Number	Location ID code
ParkCode	Text	4-character Park Code
Project	Text	Code for component of program (Weather, Birds, Fish, Veg Plots, etc.)
Description	Text	Description of this location (<256 characters)
StartUTMX	Number	UTM X (easting) coordinate for the center of the plot or location OR starting point of a line or polygon.
StartUTMY	Number	UTM Y (northing) coordinate for the center of the plot or location OR starting point of a line or polygon.
StopUTMX	Number	UTM X coordinate (easting) of ending point of line or polygon (double precision to15 significant digits)
StopUTMY	Number	UTM Y coordinate (northing) of ending point of line or polygon (double precision to15 significant digits)
UtmZone	Text	UTM zone
StartLat	Number	Latitude in decimal degrees for the center of the plot or location OR starting point of a line or polygon.
StartLon	Number	Longitude in decimal degrees for the center of the plot or location OR starting point of a line or polygon.
StopLat	Number	Latitude in decimal degrees for the ending point of a line or polygon.
StopLon	Number	Longitude in decimal degrees for the ending point of a line or polygon.
Datum	Text	Datum of mapping ellipsoid
EstHError	Number	Estimated horizontal accuracy error--see users guide for complete details and examples
AccNotes	Memo	Comments about how positional (horizontal) accuracy was estimated
Elevation	Number	Elevation in meters
EstVError	Number	Estimated vertical accuracy error in meters
Aspect	Text	Aspect (N, S, E, W)
Slope	Number	Slope in degrees
Directions	Memo	Directions to access site
Comments	Text	

Location ID: Unique identifying code for the location of data collection. Text field only. Embedding the project code in the LocationID will prevent duplicate LocationID values between projects occurring in similar locations.

ParkCode: Set default value to your so that manual data entry is not required.

ProjectCode: Set default value to your project code so that manual data entry is not required.

StartUTMX/Y: UTM coordinates of the site of data collection, the center point of a polygon representing site of data collection, or the start point of a line or polygon representing site of data collection.

StopUTMX/Y: UTM coordinates of the end point of a line or polygon representing site of data collection.

UtmZone: Zone 10

StartLat/Lon: StartUTMX/Y values converted to latitude and longitude.

StopLat/Lon: StopUTMX/Y values converted to latitude and longitude.

Datum: NAD83

EstHError: Estimated horizontal error in meters obtained from GPS unit.

EstVError: Estimated horizontal error in meters obtained from GPS unit.

Common Lookup Tables

1. **Parks Table (tluParkCode)**

tluParkCode : Table

Field Name	Data Type	Description
PARKCODE	Text	4-character park code.
PARKNAME	Text	Park name.
PARKTYPE	Text	Park type (National Park, National Monument, etc.).
REGION	Text	NPS Region.

2. **Projects Table (tluProject)**

tluProject : Table

Field Name	Data Type	Description
ProjectCode	Text	Code for the monitoring project (up to 10 characters).
ProjectTitle	Text	Full project title.
ProjectManager	Text	Individual in charge of monitoring project.
StartDate	Date/Time	When monitoring project was initiated.
Comments	Text	Comments on the project.
DatabaseName	Text	Name of the data warehouse.
Format	Text	Type of database.
DatabasePath	Text	Directory path of database.
DataTypes	Text	Indicates types ofdata stored in the database.

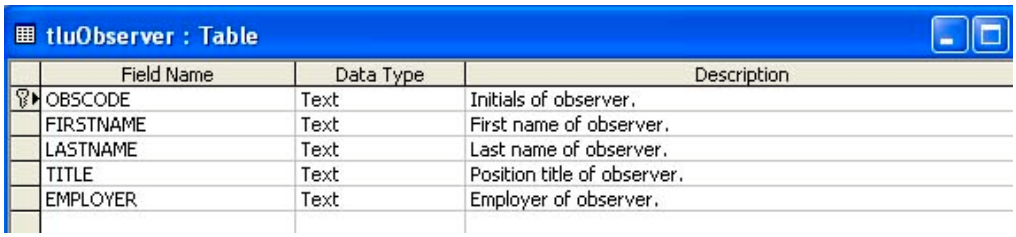
Projects table populated and maintained by Data Manager.

3. **Species Table (tluSpecies)**

Table design in development. Species table populated and maintained by Data Manager.

Project Specific Lookup Tables

1. **Observers Tables (tluObservers)**



tluObserver : Table

Field Name	Data Type	Description
OBSCODE	Text	Initials of observer.
FIRSTNAME	Text	First name of observer.
LASTNAME	Text	Last name of observer.
TITLE	Text	Position title of observer.
EMPLOYER	Text	Employer of observer.

OBSCODE: Initials of observer. Unique text field.

TITLE: Include volunteers.

EMPLOYER: Include volunteers.

Optional: Optional data may include contact information.

ADDITIONAL DATABASE DESIGN CONSIDERATIONS

Object Names

The Events, Locations, Parks, Projects, and Observers tables should maintain the same names as described above. Other table names should begin with the code “tbl”, lookup table names with “tlu”, form names with “frm”, query names with “qry”, report names with “rpt”, and macro names with “mcr”. Object names may not contain spaces.

Validation Rules / Combo Boxes

Wherever possible, validation rules or combo boxes should be employed in order to improve the quality of data entry. For number fields, validation rules can limit the range of values acceptable. Combo boxes require that the user select a text value from a list of acceptable values.

Field Names and Descriptions

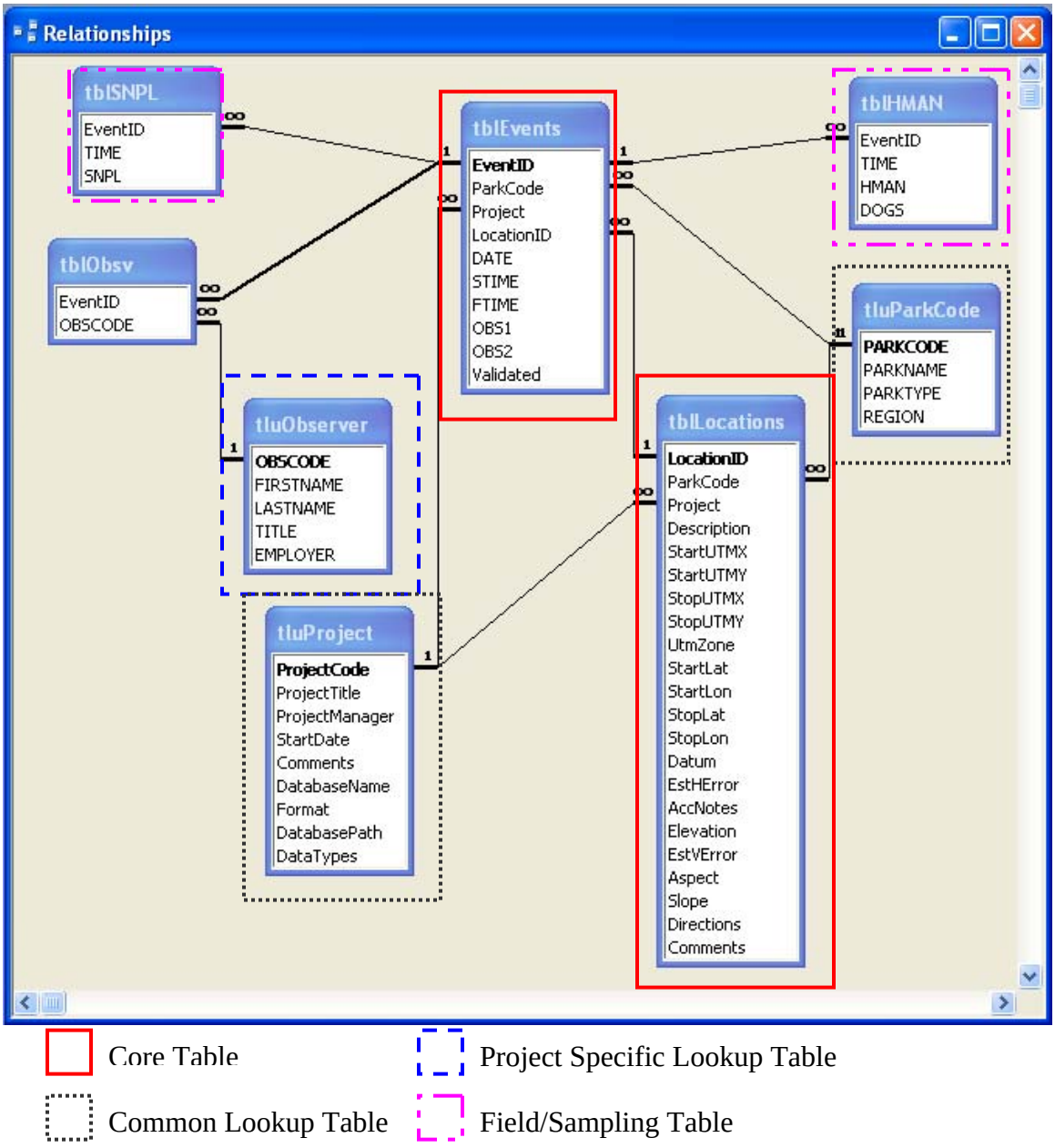
Field names may not contain spaces. A description or definition of each field is required in the table design view. Include units for all number fields.

Date and Time Fields

Date and time fields will be set as “Date/Time” data types in the table design. Date fields should be set to a short date format (ie. 9/28/2004). Time formats should be set to a short time format (ie. 15:30), with time values reported in 24hr time.

DATABASE RELATIONSHIPS

The following example illustrates a complete designation of relationships between the required tables in a standard SFAN I&M database. The example database represents data collected on the numbers of western snowy plovers, humans, and dogs observed during beach surveys.



Appendix D. SFAN Inventory Projects

Inventory Project by Type	Park
Vascular Plants (5)	
Herbarium assessment	ALL
Rare plant	GOGA
Rare plant	PORE
Vascular plant, new lands	PINN
Vascular plant (veg inventory)	EUON / JOMU
Vertebrates (11)	
Nearshore fish (coastal biophy)	GOGA / PORE
Tidewater goby	GOGA / PORE
Ashy storm-petrel	GOGA / PORE
Landbird	EUON / JOMU
Landbird	PINN
Bat	EUON / JOMU / PORE
Bat	GOGA
Bat	PINN
Mouse (salt marsh, jumping)	GOGA
Small mammal/ herpetofauna	PINN
Terrestrial vertebrate multi-spp	EUON / GOGA / JOMU / PORE
Habitat/multiple species (3)	
Coastal biological resources	GOGA / PORE
Riparian aquatic species	PINN
Sub-tidal/deep water resources	GOGA / PORE
Vegetation mapping (5)	
Vegetation map	FOPO / GOGA / MUWO / PORE
Vegetation map	JOMU
Vegetation map	PINN
Wetland map	GOGA
Wetland map	PINN
Non-vascular plants (1)	
Lichens	PINN
Invertebrates (3)	
Butterfly/moth (Invertebrates)	JOMU / PINN
Hymenoptera (Invertebrates)	JOMU
California freshwater shrimp	GOGA / PORE
Abiotic surveys (2)	
Geomorphology	JOMU
Soil survey scoping	PINN

Appendix E. SFAN Vital Sign Monitoring Programs

Vital Signs indicators selected by the SFAN as high priority for protocol development and placed within the national vital signs framework categories.

Level 1 Category	Level 2 Category	Level 3 Category	Network Vital Sign	Rank
Air and Climate	Air Quality	Ozone	Ozone levels	4
		Wet and dry deposition	Wet and Dry Deposition	4
		Visibility and particulate matter	Visibility and particulate matter	4
		Air contaminants	Air contaminants	4
	Weather and Climate	Weather and Climate	Weather/Climate	1
Water	Hydrology	Surface water dynamics	Freshwater Dynamics	14
	Water Quality	Water chemistry	Freshwater Quality	3
Biological Integrity	Invasive Species	Invasive/Exotic plants	Invasive Plant Species (early detection)	2
	Focal Species or Communities	Wetland communities	Wetlands	15
		Riparian communities	Riparian Habitat	16
		Fish	Stream T&E Species & Fish Assemblages (Salmonids)	5
		Amphibians and Reptiles	Amphibians and Reptiles	8
		Birds	Landbirds	17
		Birds	Raptors and Condors	18
		Mammals	Pinnipeds	10
		Vegetation communities	Plant Community Change (at two different scales)	11
	At-risk Biota	T&E species and communities	Northern Spotted Owl	7
		T&E species and communities	Rare, T&E Plant Species	6
		T&E species and communities	T&E Butterflies	13
		T&E species and communities	Western Snowy Plover	9
Ecosystem Pattern and Processes	Land Cover and Use	Land cover and use	Landscape & Land Use Change	12

Appendix F. SFAN Vital Sign Monitoring Program Schedules

Level 1	Level 2	Level 3	Network Vital Sign Name	Data Collection	Data Entry, Verification and Validation	Data Analysis / Reporting	Person Responsible for Data Analysis & Reporting	Type of Data Analysis	Data Archival	Long-term Trends Reporting
Air and Climate	Air Quality	Ozone	Ozone	Continuous	Continuous	Continuous	ARD	Summary statistics	Continuous	10 years
		Wet and Dry Deposition	Wet and Dry Deposition	Continuous	Continuous	Continuous	ARD	Summary statistics	Continuous	10 years
		Visibility and Particulate Matter	Visibility and Particulate Matter	Continuous	Continuous	Continuous	ARD	Summary statistics	Continuous	10 years
		Air Contaminants	Air Contaminants	Continuous	Continuous	Continuous	ARD	Summary statistics	Continuous	10 years
	Weather and Climate	Weather and Climate	Weather and Climate	Continuous	Continuous	Continuous	Physical Science Tech, GOGA Hydrologist	Summary statistics	Continuous	10 years
Water	Hydrology	Surface water dynamics	Freshwater Dynamics	Continuous, monthly, and storm-based	Continuous	Annually	Network Physical Science Technician, GOGA Hydrologist	Hydrographs, rating curves, other hydrologic summaries	Annual	5 years
	Water Quality	Water Chemistry	Freshwater Quality	Continuous, Monthly	Continuous	Annually	Network Physical Scientist	Summary statistics, seasonal, trends	Annual	5 years
Biological Integrity	Invasive Species	Invasive/Exotic Plants	Invasive Plant Species (early detection)	Annually	Continuous	Fall/Annually	Network Ecologist/Park Biotechs	GIS analysis, summary statistics	Annual	5 years
	Fo	Wetland Communities	Wetlands	TBD	TBD	TBD	TBD	TBD	TBD	TBD
		Riparian Communities	Riparian Habitat	TBD	TBD	TBD	TBD	TBD	TBD	TBD

Level 1	Level 2	Level 3	Network Vital Sign Name	Data Collection	Data Entry, Verification and Validation	Data Analysis / Reporting	Person Responsible for Data Analysis & Reporting	Type of Data Analysis	Data Archival	Long-term Trends Reporting
		Fishes	Stream Fish Assemblages	Annually	Annually	Annually	PORE Hydrologist/ Fish Tech.	Population estimates	Annual	3 years
		Amphibians and Reptiles	Amphibians and Reptiles	TBD	TBD	TBD	TBD	TBD	TBD	TBD
		Birds	Landbirds	Annually	Annually	Annually	PRBO	Summary statistics, population trends	Annually	5 years
		Birds	Raptors and Condors	Annually	Annually	August 30, annual	Wildlife tech, PINN Biologist	Nesting success, locations, summary statistics	Annually	5 years
		Mammals	Pinnipeds	Bi-weekly/ Annual	Weekly	August 30, annual	BioTech, PORE I&M Coordinator	Summary statistics	Annual	5 years
		Vegetation Communities	Plant Community Change (at two different scales)	For 3 years every 10 years	Continuous during collection	Year 4	PORE GIS Biologist/ Plant Comm. Ecologist	TBD	Year 4	20 years
	At-risk Biota	T&E Species and Communities	Northern Spotted Owl	Mar 15-July30/ Annual	Julyl 5-Aug 15	August 30, annual	BioTech, PORE I&M Coordinator	Summary statistics	Annual	5 years
		T&E Species and Communities	Rare, T&E Plant Species	1 Sep – 15 Oct/ Cyclic	Continuous during collection	Fall/ Annual	Park Plant Ecologists/ Biotechs	Summary statistics, population trends	Annual	3 – 5 years
		T&E Species and Communities	T&E Butterflies	Mar 15 – Sept 30/ Cyclic	Continuous during collection	Fall / Annual	PORE I&M Coordinator/ GOGA Wildlife Ecologist	Nectar source density trends, relative population estimates	Fall/ Annual	6-8 years
		T&E Species and Communities	Western Snowy Plover	Bi-weekly and Bi-monthly/ Year round	Continuous during collection	Fall/ Annual	PRBO/ PORE I&M Coordinator/ GOGA Wildlife Ecologist	Population and productivity estimates, summary statistics	Fall/ Annual	5 years

Level 1	Level 2	Level 3	Network Vital Sign Name	Data Collection	Data Entry, Verification and Validation	Data Analysis / Reporting	Person Responsible for Data Analysis & Reporting	Type of Data Analysis	Data Archival	Long-term Trends Reporting
Ecosystem Patterns and Processes	Land Cover and Use	Land Cover and Use	Landscape and Land Use Change	TBD	TBD	TBD	TBD	TBD	NA	10 years